AD NUMBER	
AD392886	
CLASSIFICATION CHANGES	
TO:	unclassified
FROM:	confidential
LIMITATION CHANGES	
TO:	Approved for public release; distribution is unlimited.
FROM:	Distribution authorized to DoD only; Administrative/Operational Use; SEP 1968. Other requests shall be referred to Office Chief of Research and Development, Attn: STA Division, (Army), Washington, DC 20310.
AUTHORITY	
30 Sep 1980, Group-3, DoDD 5200.10, per document marking; DAMA ltr dtd 13 May 1975	

Cop. 4

[redacted]

TECHNICAL PAPER
RAC-TP-309

SEPTEMBER 1968

RESEARCH ANALYSIS CORPORATION

Measurement of Pacification Progress in Vietnam (U)

by
James W. Johnson
Charles Anello

PRO CON
RAC

Classified material appears only on pages 22, 27, 33–36, 51–59, and 64–66. Classification of paragraphs on these pages is as indicated.

DISTRIBUTION STATEMENT
In addition to security requirements which apply to this document and must be met, each transmittal outside the Department of Defense must have prior approval of Office of Chief of Research and Development, ATTN: STA Division, Department of the Army, Washington, D. C., 20310.

PREPARED FOR THE DEPARTMENT OF THE ARMY
Contract No. DA 44-188-ARO-1

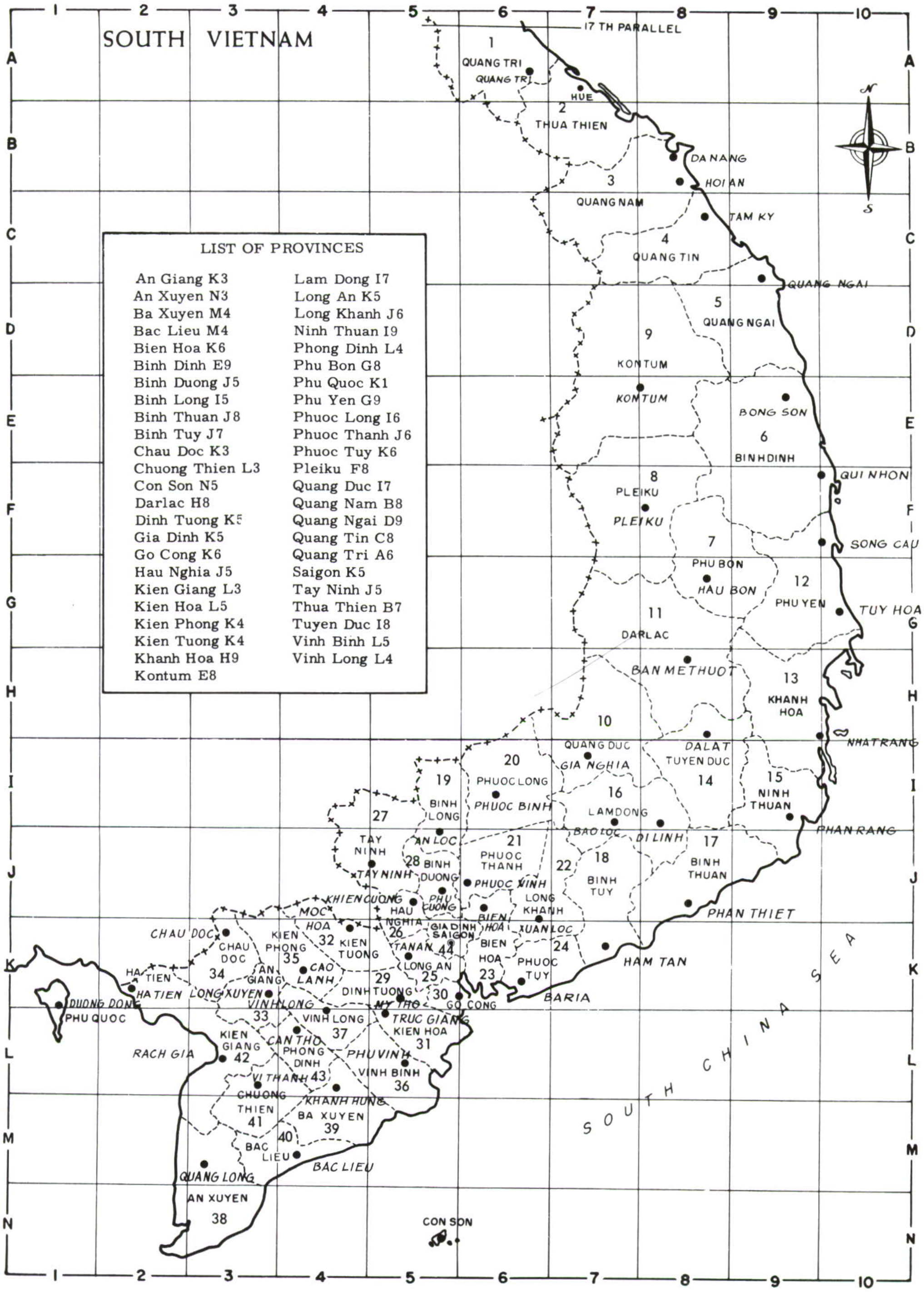
GROUP 4
DOWNGRADED AT 3-YEAR INTERVALS; DECLASSIFIED AFTER 12 YEARS

Copy 169 of 175

[redacted]

[redacted]

[redacted]



Map of South Vietnam

With province key numbers and letter and number coordinates.

[redacted]

MILITARY GAMING DEPARTMENT
TECHNICAL PAPER RAC-TP-309
Published September 1968

Measurement of Pacification Progress in Vietnam (U)

by
James W. Johnson
Charles Anello

AD-392 886

RESEARCH ANALYSIS CORPORATION
McLEAN, VIRGINIA

[redacted]

DEPARTMENT OF THE ARMY
OFFICE OF THE CHIEF OF RESEARCH AND DEVELOPMENT
WASHINGTON, D.C. 20310

CRDSTA

This document was prepared by the Research Analysis Corporation under an institutional research program. Prior to the implementation of the Hamlet Evaluation System (HES) in South Vietnam in January 1967, the collection and organization of pacification information was extremely difficult. Considerable uncertainty existed as to which variables represented meaningful gauges of pacification and what overall measure was suitable as a measure of progress. Within this context, RAC-TP-309 appears to represent a pioneer effort in this area.

The multiple regression technique employed in RAC-TP-309 is one of the multivariate techniques being applied to evaluate current pacification data in South Vietnam.

RUSSELL D. McGOVERN
Colonel, GS
Chief, Studies and Analyses
Division

Published September 1968
by
RESEARCH ANALYSIS CORPORATION
McLean, Virginia 22101

FOREWORD

By far the largest proportion of time and effort expended on military operations research and systems analysis has been concerned with past history and the hypothetical future. Further, the major efforts have been allocated to large-scale conventional and nuclear warfare. The challenge of the moment is to deal concurrently with the real-world real-time conflict in Vietnam and elsewhere in the world where kindred problems exist or may erupt. This means analysis of forms of limited warfare, actual or potential, which also include directly a significant proportion of political, economic, psychological, and sociological considerations.

It follows that requirements exist for methods or analytical models that are suitable and adequate for gaining greater understanding of insurgency/counterinsurgency-type situations; for measuring progress; for detecting and assessing causal relations; and for evaluating alternative allocations of effort, tactics, and strategies.

This paper reports an exploratory investigation generated by the challenge and the requirements. The extent to which the steps taken and the results achieved merit further effort depends on subsequent professional evaluation. Comments are solicited.

Lawrence J. Dondero
Head, Military Gaming Department

CONTENTS

Appendixes

Figure

Tables

Problem

To develop and test a mathematical model for the quantitative measurement of pacification progress in Vietnam, to identify and measure factors affecting progress, to develop a useful predictive device, and to provide a method for checking the reliability of the data-reporting systems in the field.

Facts

US involvement in Vietnam depends ultimately on judgments concerning US national interests and national strategy. The current phase of US commitment began in 1954 with the provision of advice on military matters, military equipment, training, and economic support. The level of support has gradually increased. By 1962 the US advisory and limited direct military assistance was increased when it appeared that the Republic of Vietnam (RVN) efforts were succeeding against the Viet Cong (VC) and could be reinforced advantageously. However, after the Diem coup d'état in November 1963, the rapid gain of the VC (with increasing outside sponsorship) culminated in a US decision to introduce a large number of combat and support forces and provide additional economic assistance.

Throughout, decisions depended in part on the measurement of RVN pacification progress, or conversely, VC success. Such measurement, of course, takes into account simultaneously many aspects of the situation. This study is presented as a contribution to the development of methods for the measurement process.

Three levels of violence or armed conflict exist in Vietnam: terrorism, paramilitary or guerrilla actions, and limited conventional warfare. Simultaneously, and closely integrated, there is the "other war," involving political, economic, psychological, and sociological activities.

A wealth of quantitative (as well as qualitative) data has been and continues to be generated and reported on the various aspects of the conflict. Among the sources of data are the Government of South Vietnam (GVN) through its administrative machinery at all levels, starting at the hamlets; the US advisors under the US Military Assistance Command, Vietnam (USMACV); the US Embassy; the US Agency for International Development (AID); the US Central Intelligence Agency (CIA); and the US combat forces.

Data vary in quality, adequacy, and availability for analytical purposes. For this investigation the data employed were presumed to be at least adequate for exploratory purposes.

A variety of more or less competitive mathematical techniques exists for structuring analytical models. For this study the linear multiple regression technique was employed as the principal method.

Discussion

Methods

Pacification progress,* for purposes of this study, was measured by the proportion of the "population and area control" or "population control" exercised by each side.† The suitability of the mathematical model was evaluated by standard statistical tests. The precision and potential usefulness of the results of the analysis were assessed by comparing the population-control data generated in the field in Vietnam with those computed by the model. The rationale for the overall investigation follows.

The four basic elements involved in the development of an analytical model are concept, analytical techniques, data, and the selection of factors or variables relevant to the problem. The discussion of concept, or theory, is limited to those considerations essential to understanding the method.

TABLE 1

Phases of Insurgency Measured by Population and Area Control or Dominance by the Insurgent

Phase	Percent controlled[a]
Incipient	0–20
Developing	20–40
Strategic defense	40–60
Strategic offense	60–80
Climactic	80–100

[a]Scale developed by historical inference.

The first consideration was the identification of the phase of the conflict (insurgency) as developing, strategic defense, or strategic offense. A more quantitative set of phases, based on the proportion of the populace generally supporting or under the control of the insurgents, is listed in Table 1.

*The term "pacification," as used here, encompasses all military and nonmilitary actions and considerations that result in establishing and maintaining control, dominance, or effective influence over the populace and the physical environment by either side.

†Control connotes the degree of security established as well as the degree of influence exercised over the populace.

During the period investigated in this study,* VC countrywide control was estimated by the USMACV to be in the lower to middle portion of the 40 to 60 percent interval.

The second consideration was the selection of a measure for pacification progress. As the result of the province-by-province analysis, covering a range of VC control from approximately 0 to about 80 percent, population control was selected as the most useful single measure.

Since at least July 1962, population-control data have been reported at intervals in Vietnam and, since May 1964, monthly by province, thus providing a reasonably consistent base line for analysis in spite of modest definitional changes. At least to a first approximation, the criteria employed for estimating population and area control are also used to evaluate individual and community security and communications and resources control. Population control, in addition, reflects the strengths and effectiveness of the opposing forces as well as nonmilitary civic-action efforts.

Five distinct categories of population control are reported in Vietnam. "Secured" means essentially 100 percent control by the RVN. "Undergoing Securing" means that the RVN has achieved dominance, including virtual elimination of VC principal combat elements, and civic-action efforts are under way to gain positive allegiance of the populace and thus to establish a Secured status. "Undergoing Clearing" refers to areas (and the included populace) still contested, with friendly forces more or less active in eliminating VC combat units. On balance, the VC remain dominant. The category "Uncontested" includes areas considered to have little or no tactical or strategic value for either side; they are either very lightly populated or uninhabitable. "VC Controlled" is the converse of RVN Secured, i.e., essentially 100 percent control by the insurgent.

For purposes of analysis an examination of each of these categories is useful. Secured and VC Controlled are essentially unambiguous. Undergoing Securing generally reflects the level of effort being expended by civil-pacification agencies. To an extent, Undergoing Clearing measures the level and effectiveness of the major military efforts. However, it was considered worth while also to combine the five categories into a single value by weighting factors based on the definitions. Table 2 exhibits the results in percentages of each category to be distributed to either RVN or VC control.

The category Uncontested, involving less than 1 percent of the total population, was eliminated, and the values were transferred to Undergoing Clearing or VC Controlled, as appropriate. Although this single-value approach should be viewed as somewhat experimental, the results of the analysis appear to confirm a high degree of reliability.

*A 22-month test period, May 1964—February 1966, was chosen to coincide with the initiation of a new reporting system and was limited by data availability. Should the investigation be extended, the data limitations can be removed.

The third principal consideration was the choice of factors (from a list of over a hundred candidates) that can be presumed to relate to or affect pacification progress and, more specifically, population control. The choice is limited by data availability and the desirability of using the fewest essential factors in the analytical model.

TABLE 2

Weighting Factors for Population and Area Control

Categories reported	RVN control weights, %	VC control weights, %
Secured	100	0
Undergoing Securing	70	30
Undergoing Clearing	40	60
Uncontested	0	0
VC Controlled	0	100

Conceptually, factor selection also depends on the phase of insurgency. For the period tested, the primary general consideration affecting control appeared to be security, and for security the military factors would be dominant. The factors selected would thus include all types of opposing forces engaged and all activities directed primarily against military targets. This hypothesis appears to have been supported by the results obtained. To forestall any possible misunderstanding, it is fully recognized that other factors—political, economic, psychological, sociological, demographic, etc—no doubt affect control. This would be especially true, for example, in the incipient insurgency phase. Any subsequent refinement of the model would test for the effects of additional factor candidates.

The mathematical model employed is the standard linear multiple regression technique. Population control is defined as the "dependent variable," and the factors that relate to or that are presumed to affect control are called the "independent variables."

Starting with a concept that expresses pacification in terms of the proportion of the population subject to VC control (denoted by Y) and a set of selected variables apparently related to control (denoted X_i, $i = 1, \ldots, r$) such as number of RVN and VC forces, the problem is to select a mathematical form that will relate control to these independently reported variables. In this instance it was desired to learn whether the available data exhibited enough consistency that control could be approximated by a linear combination of these related variables and, more specifically, whether control could be represented by the relation

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r$$

where β_0 is the underlying level of control in a province, and β_i, $i = 1, \ldots, r$,

is the change in control due to changes in the ith independent variable. This model is called a linear multiple regression model. It is linear in the β's.

Results

The results are based on computations of population control by the linear multiple regression model for each of the 43 provinces in Vietnam over a 22-month test period May 1964–February 1966. Especially relevant results are listed as follows:

(a) With the possible exception of a few provinces, the model could have been used as an alternative method to the field-type estimates made routinely in Vietnam for measuring pacification progress.

(b) The results were sufficiently precise that the overall trends in population control, that is, the direction and rate of change (percent per unit time), could be observed reliably.

(c) Of the independent variables selected and used, those that best related to and affected population control were identified.

(d) Effects of changes in the values of the independent variables [for example, the number of RVN Regional Forces (RF) and Popular Forces (PF)] on population control can be calculated.

(e) The overall analysis, including the process of adjusting the basic data to remove the more evident inconsistencies, indicates that the field-reporting systems could have been monitored advantageously for the early detection and correction of data discrepancies.

TABLE 3

Summary of Results for All 43 Provinces

Percentage relative error	Number of provinces	Percentage of VC Controlled population, weighted included	Cumulative percentage of VC Controlled population, weighted included
≤ 2	12	52.0	52.0
≤ 3	8	24.0	76.0
≤ 5	6	8.0	84.0
≤ 10	13	13.5	97.5
> 10	4	2.5	100.0
	43		

Table 3 summarizes results for all provinces, and Table 4 for the whole country and the four corps tactical zones. In order to take into account, simultaneously, all steps taken in the analysis to include data adjustment and the weighting of the five categories of control reported to produce a single value, the data in the tables are for VC Controlled population, weighted (see Table 2).

The value shown as "percentage relative error" represents the difference between the adjusted field-reported value and that computed by the model.

In Table 3 it is observed that over half the estimated populace under VC control was computed to within 2 percent; 84 percent to within 5 percent; and 97.5 percent to within about 10 percent. However, as noted in detail in later sections, these results understate the degree of precision achieved by the analysis. Of the 17 provinces with percentage errors of over 5 percent, (a) 10 included relatively small proportions of overall VC control, and thus even small computed percentage-point differences inflated the errors; (b) 11 contained relatively small and widely distributed populations, and thus minor data discrepancies again inflated the errors; (c) five were involved in significant boundary changes with the result that all categories of data had to be somewhat arbitrarily redistributed, and data-reporting discrepancies were observed during the periods when the changes took place; and finally (d) a single province, Gia Dinh, also known as the Capital Military District (CMD), presented special data problems, and the relatively less precise computations of control had a disproportionately large effect on the results.

TABLE 4

Summary of Countrywide Results, by Corps

Area	Average total population, thous	Average VC Controlled population, weighted, thous	Percentage relative error
I Corps	2,305	1234	1.1
II Corps	2,565	1223	1.3
III Corps	2,788	918	2.4
IV Corps	5,468	2401	0.9
Countrywide	13,126	5776	0.6

Table 4 exhibits the results for the whole country, and the four corps. The small values of the percentage relative errors speak for themselves. If Gia Dinh were omitted from III Corps, the percentage relative error would have been about 1.5 percent.

Conclusions

1. A quantitative method for measuring pacification progress in Vietnam, as defined in this study, has been developed and tested using the linear multiple regression technique.
2. The results, including the degree of precision achieved, appear to support the view that a more comprehensive model could be developed to measure pacification progress even more precisely, to identify and quantify the causal relations, to estimate force capabilities and requirements, and for prediction purposes.
3. The overall analytical model could be used advantageously to check and improve the various ongoing data-reporting systems in Vietnam.

Measurement of Pacification Progress in Vietnam

ABSTRACT

The problem was to develop and test a quantitative method for measuring the progress of pacification in Vietnam. Pacification, for purposes of this study, means the establishment of internal security, political stability, and economic viability and the elimination of external belligerent intervention.

Pacification progress was measured by the proportion of the population and areas controlled, dominated, or influenced by one side or the other; the rates of change in such proportions; and the direction of change. Population and area control also includes communications and resources control and reflects the strengths and effectiveness of the opposing forces.

Since July 1962, population and area-control data have been officially reported in Vietnam at intervals in several related formats. Concurrently, data are also reported on a variety of factors, such as opposing forces strengths and their activities, that may be presumed to relate to or affect control. This study investigates the feasibility, suitability, and reliability of an alternative method for measuring control by an analytical model in place of the field estimates. The model is based primarily on the standard linear multiple regression method. Computations were made by month for a test period of 22 months (May 1964–February 1966) for all 43 provinces. The province results may be combined as desired for other geographical or administrative areas such as division and corps tactical zones and the entire country.

To evaluate study results, comparisons were made of the reported and the computed levels of control, and tests were made of statistical hypotheses concerning certain aspects of the linear model. Comparisons of the reported and computed values reveal that the countrywide difference was less than 1 percent, the difference in corps tactical zones less than 2 percent, and the difference by provinces less than 5 percent in the majority of cases. In retrospect, readily observable basic data deficiencies, for which corrective measures could be taken, account for most of the larger differences.

The evidence appears to support the conclusion that the field-reported control data exhibit enough consistency to permit using an alternative computational technique based on a linear model instead of field estimates of control. This independent measure of pacification progress is shown to be useful for testing and evaluating the reporting systems in Vietnam. The principal recommendation is to extend the investigation for purposes of more precisely identifying and evaluating the factors affecting pacification progress, estimating force capabilities and requirements, and predicting future trends.

ABBREVIATIONS

ACY	Armed Combat Youth (Republic of Vietnam)
ARVN	Army of the Republic of Vietnam
CIA	Central Intelligence Agency
CIDG	Civilian Irregular Defense Group (Republic of Vietnam)
CMD	Capital Military District (Republic of Vietnam)
d_k	population density deviation index
GVN	Government of South Vietnam
NVA	North Vietnamese Army
PF	Popular Forces (Republic of Vietnam)
RF	Regional Forces (Republic of Vietnam)
RVN	Republic of Vietnam (South Vietnam)
USAID	US Agency for International Development
USMACV	US Military Assistance Command, Vietnam
VC	Viet Cong

INTRODUCTION

The basic reasons for US involvement in Southeast Asia, especially the current involvement in Vietnam, are the US national interests and the national strategy. Within South Vietnam the US objective is to assist in establishing an environment of political stability, economic viability, and military security against internal insurgency and external belligerent intervention.

Until 1954 such involvement was largely indirect support of French efforts to regain and maintain its colonial status in Indochina. In 1954, with the abdication of the French position after Dien Bien Phu, the US almost unilaterally assumed the burden of helping to establish a semblance of stability in the south.

During the 1961–1963 period the threat against the Diem regime in Saigon increased after the establishment of the VC National Liberation Front. By 1962 Diem's Strategic Hamlet Program promised a substantial counter to this threat. The US gradually increased its economic and military aid and enlarged its advisory mission.

By September 1963 it appeared that progress was such that the US could withdraw completely by the end of 1965. However, after the Diem coup d'état in November 1963, the weaknesses of the GVN counterinsurgency program were revealed, and a high degree of internal political instability ensued.

By mid-1964 it became increasingly apparent that the VC was improving its posture significantly, for example, its increasing strength in the provinces encircling Saigon. Later in 1964 a VC buildup in the highland region (II Corps) threatened to cut South Vietnam in two. Reports of North Vietnam's infiltration of regular forces increased. By March 1965 the US increased its commitment by introducing combat forces into South Vietnam. Subsequently the size of the US (and allied) forces reached about half a million men.

The situation in Vietnam and the increasing US involvement has been reviewed here simply to observe that the assessment of the situation has been a continuing process. Many considerations were involved in the decision-making process. One of the elements in the estimate of the situation in Vietnam has been the measurement of pacification progress.

Since about mid-1962, one method employed routinely to analyze or measure pacification progress in Vietnam was to estimate the proportion of the populace under the control, dominance, or effective influence of one side or the other. The principal objective of this study was to explore the feasibility and suitability of a mathematical model for the quantitative measurement of pacification progress. The standard linear multiple regression method was selected as the principal technique.

Other objectives included the identification and measurement of the factors affecting progress and the development of a useful predictive device. Further, it was envisaged that a considerably more comprehensive analytical model, capable of more precise analysis of pacification progress, estimation of force capabilities and requirements, and identification of causal factors and useful for both short- and long-range forecasts could be developed in due course, employing this study as a point of departure.

It was also anticipated that the investigation could provide a method for checking and possibly improving the reliability of the data-reporting systems in the field.

The extent to which these objectives were achieved is determined by standard statistical tests of the model and by analysis of the results of a comparison between the officially reported data and those computed by the mathematical model.

BACKGROUND

The following background discussion of the situation in Vietnam is necessary in order to develop a concept of the conflict for analytical purposes. It is limited to considerations affecting the selection of the analytical model, the selection of the factor or factors to be measured, and the related variables.

It is generally recognized that the conflict includes political, economic, sociological, and psychological considerations as well as military and police actions. The effects of historical values in any given country or society play a major but not readily calculable role; in Vietnam, for instance, the feeling for independence is strong in spite of or because of various degrees of subservience to foreign domination, most recently to that of France as part of its colonial empire. Outside sponsor influences, including the communist theory of "wars of national liberation" and the direct intrusion by a communist regime of North Vietnam (supported by the USSR and Communist China) into the south to aid and abet the insurgents, are important determinants of the course of the insurgency.

In South Vietnam local problems such as relative dominance of the political scene by individuals of North Vietnamese origin; endemic political instability; the existence of ethnic minorities such as the Chams, the Chinese, the Montagnards in the central highlands, and the Khmers in the delta; and religious differences between the Catholics and Buddhists, as well as the Hoa Hao and Cao Dai sects, affect allegiances and attitudes of the people.

The geography and demography of the country play significant roles in understanding and analyzing the conflict. For example, two political power centers exist in the RVN, oriented toward Hué in the north and Saigon in the south. The geographical divisions of the country—the coastal lowland, the highland and plateau, and the delta regions—differ in terrain, food production, land distribution, and other societal patterns. The size, distribution, and density of the population differ markedly within the various regions and among the districts and provinces. Political, economic, and sociological outlooks differ between the city and rural populace. In June 1967, of the estimated total population of 16.5 million, over 2.7 million reside in the six autonomous cities:

Saigon-Cholon, Hué, Danang, Dalat, Vung Tau, and Cam Ranh. The rural populace of about 14 million (increasing about 2 percent annually) are unevenly distributed among 43 or 44 provinces (depending on the recent time period selected), about 240 districts, 2200 villages,* and some 12,300 hamlets.†

The six autonomous cities are generally regarded as being under at least nominal control of the GVN. As is typical of insurgency situations in many underdeveloped countries, the rural population is the prime target of the VC. The political, economic, and sociological horizons of the average rural Vietnamese center on his hamlet and perhaps his village, but for only about 1 percent of the total rural population do these horizons extend to the center of government in Saigon.

The decision to concentrate analysis on the rural population was based on the consideration of a variety of basic factors that influence the outcome of the conflict, the heterogeneity of the country, and the relatively neutral political attitude of the Vietnamese peasant.

Pacification

From the point of view of the US, the (assumed) immediate objective is to assist in establishing an environment in which the South Vietnamese can determine their own future destiny, with minimal internal disruptions and without external interference. For purposes of this study, the overall objective of counterinsurgency efforts is termed "pacification." The term "revolutionary development" is used somewhat synonymously, although this term is more specifically applicable to nonmilitary activities.

Pacification is divided into two general categories of activities: "the war" and "the other war." The war involves military actions principally to defeat forces infiltrated from North Vietnam and the main and local regular forces of the VC. It also means the defeat and elimination of VC paramilitary forces (guerrillas). (The application of air-delivered force by the US, and to a much lesser extent by the GVN, against North Vietnam is a part of this war; however, only activities within South Vietnam are accounted for in this analysis.)

The other war concerns the extension of GVN authority throughout the country by police actions, including those against the part-time guerrilla forces and the VC infrastructure, and by political, economic, psychological, and sociological measures intended to gain the acceptance of this GVN authority by the populace.

Phases of Insurgency

The extent to which pacification involves both wars depends on the phase of the insurgency. The communists generally speak of three phases: development, the strategic defense, and the strategic offense. It is sometimes convenient to divide the successful insurgency into five phases (as shown in Table 1)

*The number of hamlets per village range from 1 to 25, with an average of less than 10.

†Population and political-districting data are in a state of flux due to the relatively underdeveloped state of the country, lack of a complete recent census, population movements resulting from the war, the relative inaccessibility of VC Controlled areas, political boundary changes, and destruction of hamlets.

on the basis of the proportion of the populace generally supporting or under the control of the insurgent; the phases are categorized by intervals of 20 percent of the population, 0 to 100 percent. That is, 0 to 20 percent would mean insurgency in an incipient stage or phase; 20 to 40 percent, development; strategic defense, 40 to 60 percent; mobile or strategic offense, 60 to 80 percent; and a final, climactic, successful insurgency, over 80 percent control. In a typical 40 to 60 percent interval of control the insurgent would undertake high-intensity regular-force actions to defeat the counterinsurgent as he did in North Vietnam at the time of Dien Bien Phu.

The phase of insurgency may also be indicated by the level of insurgent activities, i.e., by the relative numbers and types and effectiveness of incidents perpetrated such as attacks by size and type of target, harassing fires, harassments, sabotage, assassinations and kidnappings, antiaircraft fires, and armed propaganda incidents. The total number, projected in part on historical events as in Malaya, may be expected to peak in the interval of 40 to 60 percent insurgent control, assuming that, at the same time, the counterinsurgent is not demoralized and is using all forces available to stem the tide. Such an approximate situation appeared to have developed about mid- to late-1964 and resulted in the introduction of US combat forces in strength beginning in March 1965.

POPULATION CONTROL

Selection of Criteria

Analytically the identification and selection of criteria for measuring the state of insurgency/counterinsurgency involve many factors, not all of which are directly and readily quantifiable. A list of about 100 indicators appears in Table 5.

To limit the problem to manageable proportions these were first reduced to six "yardsticks": population control, area control, communications control, resources control, numbers and effectiveness of the insurgent forces, and numbers and effectiveness of the counterinsurgent forces. Such a selection is not meant to exclude other potentially important factors listed in Table 5: casualties, defections and desertions, weapons lost and recovered, and various economic factors—tax collections, agricultural and manufacturing productivity, and the cost of living. An important constraint on criteria selection at this juncture is the availability of reliable data, especially in the form and of the consistency necessary.

Certain decisions were made at the outset of this investigation. First, the basic "echelon" to be considered would be the individual province. The elements of the provinces—population, population distribution and densities, and number of districts, villages, and hamlets—would be taken into account to the extent possible. At the same time, higher echelons would be considered: special zones, division tactical zones, and corps tactical zones as well as VC military regions. Second, mainly because of the data initially available, two measurable factors related to progress were selected: population control and VC activities. The second decision involved eliminating area, communications, and resources control per se, again in large part because of lack of

TABLE 5

Criteria for Selection of Independent Variable Factors

(For analysis of pacification progress)

Category	Type of analysis	Measurement factor	Specific objective
A	General	Time	Date, month, year
		Areas of interest	Hamlet, village, district, province, special zone, division, corps, countrywide
		Geographical regions	Coastal, plateau, highland, delta
		Demographic regions	Ethnic, minorities
		Population density	Population density deviation index d_k
B	Control pacification criteria	Population and area control	Secured, Undergoing Securing, Undergoing Clearing, Uncontested, VC Control, unweighted
		Population and area control	Weighted for preceding categories
		Area control	By areas of interest
		Communications control	Roads, railroads, rivers, canals, coastal waters, air, infiltration routes
		Resources control	Population-recruitment base, fertile land, rice, charcoal, etc
		VC activities control	See Category D
C	Order of battle	Friendly order of battle	RVN Army, Navy, Air Force, Airborne, Rangers
		Friendly order of battle	RF, PF, CIDG, National Police, Armed Combat Youth
		Friendly order of battle	US and Free World forces, ground and air
		Enemy order of battle	NVArmy, main and local forces
		Enemy order of battle	Guerrillas, self-defense, secret-self-defense, political cadre, support
D	Operations	Friendly operations	Large ground, small ground with and without contact; by types, search and destroy, search and clear, reconnaissance, etc; air, in South Vietnam and North Vietnam
		VC activities	Military and nonmilitary; attacks, harassing fires, mines, sabotage, harassments, AA fires, propaganda, kidnappings, assassinations, etc
E	Readiness, friendly and enemy	Operational readiness	Army Navy, Air Force, other
		Training and schooling	All forces
		Equipment	Readiness requirements
		Intelligence	Military, nonmilitary
		Recruitments	Local, external
		Infiltration	—
		Bases	Internal, external
F	Political, economic, psychological, sociological	Political	Elections, GVN stability, demonstrations, etc
		Economic	Cost of living, tax collections, landownership distribution, currency exchange rates, availability of goods, black market, basic crop cultivation and distribution, etc; self-help programs, AID programs; hamlet development; hamlet construction, etc

TABLE 5 (continued)

Category	Type of analysis	Measurement factor	Specific objective
		Psychological	News media, hamlet communications, voluntary intelligence provided, etc
		Sociological	Orientation of religious groups, voluntary and involuntary population movements, astrology, "Chieu Hoi" Program for Returnees, Revolutionary Development programs, integration of minorities or special communities as Hoa Hao, Cao Dai, Montagnards, Khmers, etc
G	Miscellaneous	Defections	Enemy, rate and locale
		Desertions	Friendly, military, administrative personnel
		Captured	Friendly and enemy
		Weapons movements	Lost and recovered
		Casualties	WIA, KIA, MIA; ratios between and among friendly and enemy forces by type; civilian, all categories

consistent data; also, for this investigation, the VC activities factor was eliminated as a direct measure of progress. Thus population control is the single measure of pacification progress employed in the study.

Population control was assumed to cover the other control yardsticks reasonably well. In practice, population control is calculated in the field by delineating areas and counting or estimating the number of people within them. Population control, by virtue of the definition, includes also US and Vietnamese estimates of area, communications, and resources control and reflects the strength and effectiveness of the VC.

Population control began to be reported about July 1962 by the Office of the Assistant Chief of Staff, Intelligence, J2, USMACV, at about 6-month intervals. In May 1964 a new reporting system was inaugurated by the Office of the Deputy Chief of Staff for Operations, J3. The principal data employed in this study are from the J3 source.*[1]

US sector (province) military advisors were responsible for the monthly reports. In practice, Vietnamese hamlet, village, district, and province chiefs provided the basic data; US advisors at district or lower levels assisted. Maps were drawn, delineating areas under various categories of control; the included populace was counted or estimated; and percentages were calculated by province. After the first 4 months two attachments were submitted: one, usually under the heading "Explanation of Changes," provided some detailed explanations; the other was a breakdown of the number of hamlets in each category of control, as well as those designated as being in different planning stages. USMACV inspectors were subsequently employed to visit the provinces to supervise and improve the reporting system.

*In January 1967 a new (hamlet evaluation) system began to take form; its compatibility with the preceding ones remains to be determined. A preliminary survey indicates, however, that compatibility can be achieved.

Categories of Control

The term "control" has several simultaneous connotations. Control means literally the imposition of governmental (or insurgent) force such that the people conform to dictation. It also means compliance and, ultimately, voluntary, active allegiance. Control plus voluntary allegiance means pacification.

Population control is registered in five categories: people in areas that are Secured, Undergoing Securing or Undergoing Clearing of VC forces (also referred to as "contested areas"), Uncontested, and under VC Control. (For convenience these were initially referred to by a color code, respectively: Dark Blue, Light Blue, Green, White, and Red.)

Secured areas, by definition, had to meet the following six criteria: (a) the hamlet residents have been screened and existing VC infrastructure discovered and eliminated; (b) popular or other self-defense forces have been selected, trained, and armed; (c) an obstacle system and other fortifications have been constructed for defense against VC incursions; (d) a communications system for requesting reinforcements has been set up; (e) hamlet inhabitants have been organized into groups and assigned specific tasks for hamlet security and for improvement activities (New Life, or Revolutionary Development); and (f) a hamlet committee has been elected by secret ballot in accordance with a democratic spirit (may be waived by Montagnard hamlets if a committee has been appointed instead). Thus secured areas are those completely pacified.

Undergoing Securing areas are those cleared of VC by GVN (or other friendly) regular or paramilitary forces, and in which mobile-action cadre teams (or their equivalent) have started Revolutionary Development work in the area.

Undergoing Clearing areas are those in which North Vietnamese army units and VC main and local forces are being engaged, and the Army of Vietnam (ARVN) or other friendly regular forces are capable of achieving a cleared status. VC acts of terrorism may still occur with some frequency and VC underground cells and guerrilla forces remain in the area.

Uncontested areas are those controlled by neither side, and are not likely to become significant in the insurgency/counterinsurgency struggle. The geographical characteristics are such that the areas are largely uninhabitable and have no tactical or strategic importance for either side, including possible use as VC sanctums.

VC Controlled areas are those essentially completely dominated by the VC, including security conditions equivalent to the secured areas under GVN control, and in which VC main, local, or paramilitary forces are present. Such areas may also include VC safe havens.

POPULATION CONTROL, WEIGHTED

Certain inferences as to the quality and availability of population-control data are drawn from an examination of the five control categories, both taken separately and combined.

The areas of principal interest in Vietnam are rural and relatively underdeveloped. Much of the terrain is difficult to traverse or penetrate: the coastal area is interrupted in a way that frequently requires movement by sea; the mountainous and plateau areas are usually covered with dense tropical growth; and the delta is interlaced by rivers and canals. The last complete official census was taken in the 1930's. The VC infest a large part of the countryside, and the fighting has resulted in a breakdown in communications and considerable population movement.

Administratively, data collection in Vietnam equates to "field-generated" problems as experienced in any conflict situation and exhibits the associated uncertainties. The US advisors depend largely on Vietnamese sources to whom direction must be transmitted. Both the US advisors and their Vietnamese counterparts tend to want to show progress; thus they overstate or understate their achievements depending on which course best suits their situation. Advisors' tours in the country are short, frequently less than 1 year. There is the problem of translating generally qualitative definitions into a quantitative base for analysis.

Observations of the system in operation support the following assumptions: First, at the outset of the reporting system (May 1964), the data could be expected to vary considerably in quality and conformity from province to province; maximum error would occur during the first 3 to 6 months. Second, the system subsequently would break down at intervals because of intense VC activity, inattention of the advisors, changes of personnel, and changes in administrative boundaries to delete old or form new districts and provinces. Third, real—as against apparent—changes are likely to occur relatively slowly; large discontinuities are suspect. Finally, the general trend in control is usually more important than month-to-month changes.

The actual data adjustments involved the knowledge already mentioned, e.g., an expectation of high probable error at the beginning, especially for regions known to be VC-dominated. That portion of the reports that explained the monthly changes by province was examined in detail. In particular the appearance of the word "reevaluation" was taken as an important clue to the accuracy of previous reports and as an indication of improved reliability. Relatively radical changes were suspect. Historical perspective was employed; working backward, the most recent data provided a basis for the adjustment of past reports. Although the process involved considerable judgment, and certainly could not be omniscient, the overall results are believed to represent a marked improvement. This belief appears to be supported by the results of the multiple regression analysis.

For purposes of analysis, the degree of heterogeneity among the regions and provinces of Vietnam is noteworthy. Provinces differ in area; type of terrain; productivity; population size, distribution, and density; social and religious structure; ethnic bases; political viability and orientation; the time during which one side has been in the ascendancy; and, especially during selected time intervals, friendly and enemy intentions. No single province can be considered typical. As few as seven or eight might be considered representative, especially for a limited time period.

Long An and Binh Thuan, the two provinces for which detailed data are given in App A, should not be taken as most representative, but they are useful examples.

Data-Adjustment Method

To adjust data, the number of people in each area are counted, and simple percentages, i.e., percentage of control in each of the categories, are calculated. Progress may be measured by changes in these percentages.

Two categories are generally self-explanatory: Secured areas and VC Controlled areas can be taken generally at face value as 100 percent control for analytical purposes. However, it is noted that both sides may have agents or secret sympathizers in these areas. For example, the VC at least attempt to maintain cells in Secured areas; one estimate is about three three-man cells per 1000 population. The GVN attempts to infiltrate agents into VC Controlled areas; even though VC methods involve a good deal of coercion, including assassinations and kidnappings, when VC Controlled areas are under GVN pressure, some portion of the populace may not be counted on fully to support the VC. One estimate is that about 10 percent of the population in these two areas should be considered either to actively support or to be readily induced to support the other side. In any event the numbers are usually small and influence results in a very minor way.

The Uncontested areas, especially those in the highlands that are generally populated by the Montagnards, are analytically meaningful. The Montagnards, considered somewhat subhuman by the dominant Annamites in South Vietnam, generally keep to themselves when permitted to do so. However, the total population in Uncontested areas is estimated at less than 1 percent.

The two most difficult areas to assess are those Undergoing Securing and those Undergoing Clearing. The definitions are clear but qualitative and require interpretation. What percentages or weighting factors best describe the influence of one side or the other? It was concluded that this question could be answered only by experience and experiment, and be justified or verified by the results achieved by the model's computations.

The first step in the adjustment of data was to eliminate the data under Uncontested and include them under VC Controlled or, in a few instances, under Undergoing Clearing. Next (especially after the reporting system settled down), the numbers under Secured (RVN) were accepted, except for some month-to-month smoothing. Third, the numbers under Undergoing Securing, although frequently relatively small, were assumed to be reasonably accurate, and were subjected mainly to month-to-month smoothing. There was little choice between Undergoing Clearing and VC Controlled with respect to acceptance of the reported data; whichever one showed the greater consistency tended to be used as reported, with some leaning toward VC Controlled to be conservative. Finally, if necessary, all four categories (Uncontested eliminated) were justified against the total province population, as well as between and among the categories. For example, a decrease in Undergoing Securing should be accompanied normally by an increase in Secured; or, less likely, an increase in Undergoing Clearing. An increase in Undergoing Securing may reflect a decrease in Undergoing Clearing; and Undergoing Clearing should increase with a decrease in VC Controlled. The foregoing, ultimately based on the criteria and definitions, emphasizes the interdependency among the categories.

In general the quantities under Undergoing Securing were a measure of the effort being expended by the Revolutionary Development teams. The numbers under Undergoing Clearing reflected the amount of military effort going

on to eliminate organized VC or North Vietnamese army elements, or at least the presence of RVN or other regular forces. The numbers under Secured and VC Controlled provided the least ambiguous basis for estimating the strengths and viabilities of the two sides. An attempt to provide a better basis for analyzing these strengths and viabilities for trend analysis, or at least to provide an additional check against the reporting system, involved the development and use of weighting factors for the four categories.

Employment of Weighting Factors

The initial impetus for the development and employment of weighting factors came from (a) the question as to how it might be possible to quantify the essentially qualitative definitions associated with the control categories, (b) a readily observable difficulty in comprehending the overall control status of an area (province, special zone, corps, region, or countrywide) by examining the five control categories separately, and (c) an apparent need to assess trends in pacification progress quickly. Subsequently it was believed that a single weighted control value could assist in checking the overall reporting system.

The definitions of the categories, although comprehensive in the original form, must ultimately be interpreted. First, as in the data adjustment method, Uncontested data can be eliminated by transferral to VC Controlled or Undergoing Clearing; judgment, based on known VC concentrations (as in Zones C and D) in rural areas, suggests VC Controlled. Second, throughout the study the six autonomous cities—Saigon-Cholon, Hué, Danang, Dalat, Vung Tau, and Cam Ranh—were presumed to be at least under nominal RVN control; thus only the population data for areas outside these cities were used.

Third, the same type of question can be asked about both Secured (RVN) and VC Controlled: do the definitions mean 100 percent, i.e., complete control? Or should, say, 90 percent be used to allow for something less than perfect control or alignment of the populace?

Fourth, the remaining two categories, Undergoing Securing and Undergoing Clearing, concern areas of contention, and control is divided between the two sides. Can this division of control, based on the definitions, be stated quantitatively? That is, can the reported (and adjusted) numbers be weighted to reflect, at least on the average, what proportion of the populace are actually under RVN or VC control?

In practice several sets of potentially valid weighting factors were tested. This study presents results from the one set that thus far appears to satisfy the requirements. Orientation is toward VC Control, i.e., VC Controlled, weighted. The category VC Controlled is taken as 100 percent. The category Undergoing Clearing is taken as being under 60 percent control of the VC; and Undergoing Securing, 30 percent. Secured assumes zero VC Control (or 100 percent RVN Controlled). To recapitulate, the weighting factors with respect to the VC are: VC Controlled, 1.0; Uncontested, 0, and the numbers of people redistributed; Undergoing Clearing, 0.6; Undergoing Securing, 0.3; and Secured, 0.0. (RVN Controlled, weighted, of course is, percentagewise, 100 - VC Controlled, weighted.)

The selection of values for the weights should be a continuing matter for experiment and test. Tests thus far made indicate relative insensitivity of the

results within the limits of plus or minus 0.05 to 0.1. Any changes in the definitions of the control categories, or in the overall reporting system, will dictate reassessments. The specific details (see App D) appear to support the validity of the choice made here.

FACTORS AFFECTING POPULATION CONTROL

The overall consideration in the selection of factors affecting population control is the logic of the conflict. This was determined at the outset by the selection of pacification as the objective of the counterinsurgent, and the selection of population and area control as the specific measure of pacification progress. The analytical model then requires the identification and selection of the factors that are presumed to relate to and affect population control. In the model, population control is called the "dependent variable," and the factors that relate to or affect it are called the "independent variables."

Selection of Independent Variables

The independent variable candidates are listed in Table 5. As mentioned previously, the large number reflects the complexity of the conflict, involving political, economic, sociological, demographic, military, and police considerations. The stage of development of the analytical model, data availability, and the desirability of employing the smallest possible number of independent variables are the principal practical constraints.

As was the case in the selection of population control to be the single measure of pacification progress, the selection of the independent variables depends first on the identification of the phase of the insurgency, as the variables of significance may be expected to differ or be more or less useful with the phase. It was assumed that VC countrywide control appeared to be in the 40 to 60 percent interval (and thus in the strategic-defense phase during the period tested), that internal security was the dominant consideration, and that military forces and operations would be the prime determinants of population control.

The independent variables used are listed in Table 6. These variables, by province, were the number of friendly forces, RVN regular forces (ARVN), RVN provincial forces (Regional), RVN self-defense forces [Popular, with the Armed Combat Youth (ACY) included] and the RVN Civilian Irregular Defense Group (CIDG) forces; and the VC main and local forces. Also included were the large (battalion-sized or larger) friendly operations, small operations (less than battalion-sized) with the contact with the VC, and VC activities identified as being directed against civilian and friendly military targets.

Time was included on a monthly basis and, on trial runs of the model, a demographic factor d_k that was a measure of the density and distribution of the populace.

Province. The basic geographical element employed was the province. The number, 43 or 44, has remained relatively constant; Phuoc Thanh, for example, was incorporated into four adjacent provinces (Key Nos. 20, 22, 23, and 28 in frontispiece), and the data were adjusted for this change; a more recent change to include a new province, Sa Dec, has not been accounted for.

TABLE 6

Independent Variables Selected for Analysis of Population Control

(Status of data sets)

Computer-punched card number	Name of variable	Date[a]	Data recorded
1	Areas of interest	May 64–Dec 66	Province is basic geographical element employed
2	d_k	May 64	Calculated population-density deviation index
3	Population and area control	May 64–Nov 66	Principal measure of pacification progress
4	VC incidents: military, non-military, and terror objectives	May 64–Jun 67	Attacks, sabotage, harassing fires, etc
5–1	ARVN	May 64–Oct 66	Principal regular ground forces, maneuver elements only
5–2	RF	May 64–Jul 66	Province-oriented; data presumed reasonably accurate
5–3 & 5	PF and ACY	May 64–Jul 66	Local protection at hamlet and village level; raw data used
5–4	CIDG	May 64–Feb 66	Montagnard tribesmen; raw data used
5–6	VC/NVA	May 64–Jul 66	Regular forces; long reporting intervals and delayed confirmation of data
6	VC order of battle	—	Guerrilla, political cadre, etc, not included because of gross estimation error
7	Large operations	May 64–Feb 66	Friendly operations, battalion-sized or larger; introduced by number of occurrences by province
	Small operations	May 64–Feb 66	Friendly operations, company-sized or smaller; not differentiated because of lack of detailed data

[a]The reporting system for the principal dependent variable, population and area control, came into being in May 1964; a new, revised reporting system in January 1967 is expected to require another adjustment in due course.

Population Density Deviation Index. The d_k is a calculated population density deviation index for each province, normalized for the provinces countrywide, and also calculated for each administrative or tactical zone. Provinces, zones, and regions in Vietnam differ markedly in size and number and geographical distribution of people. It was assumed that such differences would bear on the analysis, especially when attempting to combine the provinces into larger aggregations. The d_k was not used when examining provinces separately; however, such elimination should perhaps be reexamined. The basic data were as of about 1963, six standard density intervals were used;* the rural population was of primary interest, and two sets of indexes were calculated: Mod I, containing the original raw data as reported, excluded the six principal autonomous cities; Mod II, consisting of "adjusted data," also excluded the province and district capitals. The method of calculating d_k is as follows:

*Persons per square kilometer: 0-19, 19-97, 97-193, 193-290, 290-386, and over 386.

Let c = number of density categories (A, B, C, D, E, and F)

n = number of geographical divisions (provinces, corps, divisions, and military regions)

$i = 1, 2 \ldots, c$

$k = 1, 2 \ldots, n$

Let $X_i^{(k)}$ = population in ith density interval in the kth geographical division

P_k = total population of the kth geographical division for all density intervals

X_i = total population of the ith density interval for all geographical divisions

P = total population of all density intervals for all geographical divisions

$$P_i^{(k)} = \frac{X_i^{(k)}}{P_k}$$

$$P_i = \frac{X_i}{P}$$

Then

$$d_k = \sqrt{\sum_{i=1}^{n} (P_i^{(k)} - P_i)^2}$$

The sign for d_k is determined by the following rules: If the sum of the first three proportions is greater than or equal to 0.5, d_k is negative; if the sum is less than 0.5, d_k is positive.

Population and Area Control. Population and area control is taken, in this analysis, as the principal measure of the progress of the conflict, or the pacification program. Population control depends first on an assessment of areas; if the population and area is controlled, then, to a first approximation, communications and resources are also controlled.

VC Incidents. VC incidents encompass all reported VC activities, and have been reported, by province, with apparent consistency since about November 1963. The total number of categories, with some changes in definition, has ranged between 30 and 40. These, in turn, generally fall into nine more aggregated types: attacks, sabotage, harassing fires, harassments, assassinations, kidnappings, mines, antiaircraft fires, and propaganda. Finally, VC activities may be grouped, with respect to objectives, as military and non-military. Another category currently listed in some reports is terror.

Army of the Republic of Vietnam. ARVN is the principal regular ground force. Several other regular-force categories, such as the regular air force and the naval force, have not been included, largely because of data constraints with respect to location by province. The raw data available, typical of field data with some apparent discrepancies, have been adjusted just a little to account for month-to-month and province-by-province gaps. In addition, to the extent that the raw data permitted, only the maneuver elements have been included; i.e., overhead and reserves (mostly located in the vicinity of Saigon) have not been used in a province-by-province analysis.

Regional Forces. RF data are presumed to be reasonably accurate since these forces are normally province-oriented.

(U) Popular Forces and Armed Combat Youth. PF and ACY data presented some special problems, especially during 1964. Experience with the reporting system and agencies suggested some considerable discrepancies. In addition, in early 1964 a process of nationalization of all forces was instituted, and this process included incorporation of a large part of the ACY into the PF, with only a small number of ACY remaining on the force rolls. On a best-estimate basis, using the raw data, the two were combined for this analysis, as they generally may be assumed to perform the same function of local protection at the hamlet and village levels.

(U) Civilian Irregular Defense Groups. The CIDG typically are mainly Montagnard tribesmen associated with US Special Forces elements. Except for filling a small number of apparent data gaps, the raw data were used.

(U) VC and North Vietnam Army Forces. VC/NVA forces are regular VC and North Vietnam Army elements, since the VC include only the so-called main and local forces as reported in the intelligence system. The reporting intervals were frequently quite long, month-to-month changes have been typically infrequent, and, in addition, both types of reports are subject to delayed confirmations; to the extent possible, later confirmations have been taken into account.

(C) VC Order of Battle. Overall VC order-of-battle data have not been included thus far. On-the-ground experience, as well as observations made on the reported data, indicated strongly that gross estimation errors existed. That is to say, countrywide and, to a lesser extent, corps estimates were probably useful; however, at province level the apparent estimation errors seemed too large to include the data in the analysis. The categories in question include the guerrillas, self-defense types, secret self-defense, political cadre, and support. If and when these data become available, it is reasonable to assume that they should be included as variables affecting control or useful in measuring it.

(U) Large Operations. Large operations, i.e., all types of friendly operations, battalion-sized or larger, are introduced simply by the number that occurred, by province. Historically the majority or up to 90 percent have been of the search-and-destroy type. This is a very grossly aggregated variable as used here; possible refinements would be to use certain weighting factors such as number of battalion-days for each type of operation and to differentiate with respect to total size and intensity of conflict—e.g., considering separately those of the search-and-clear and clear-and-hold types.

(U) Small Operations. Small operations consist of all types of friendly operations, company-sized or smaller, that have not been differentiated in the analysis thus far, primarily because of lack of detailed data. Further, and more important, the data set includes only those that made contact with the enemy, representing a very small percentage of the total. If and when more detailed data, by province, become available, certain obvious refinements are strongly indicated.

Evaluation of Selection

(U) The extent to which the selection of variables satisfied the requirements can best be evaluated by the results (see App D). It is noted here, especially for any future investigations, that the employment of these variables,

based on an estimate of a countrywide-phase status did ignore the obvious fact that all the provinces were not in the same state of insurgency. This fact no doubt contributed to the fact that some provinces were not as well evaluated as others. However, it can also be taken as evidence that the concept of phase orientation vis-à-vis the selection of variables was a sound procedure.

ANALYTICAL TECHNIQUE

Let an observation on a specified province within South Vietnam at a fixed point in time consist of a measurement of the proportion of that province's population subject to VC control (VC Controlled). Such an observation is denoted by Y_j where the subscript j is used to identify the month in which the observation was recorded. Let it further be assumed that the value of Y_j can be approximated using a mathematical model and knowledge of the numerical values of a set of related (independent) variables. In this instance, these related variables might include the number of friendly and VC forces in the province during the month or the number of VC-initiated activities. The numerical values of these independent variables are denoted by x_{ij} where the subscript i identifies the particular variable being considered and the subscript j indicates the month.

Linear Multiple Regression Model

In this paper a standard linear multiple regression model is adopted. The model assumes that the proportion of the population under VC control can be expressed as a linear combination of supposedly related independent variables:

$$Y_j = B_o + B_i X_{ij} + \ldots + B_r X_{rj} + \epsilon_j$$

where X_{ij} = the value of the ith independent variable for the jth month
B_o = the underlying level of control
B_i = the change in control due to the ith variable after adjusting for the remaining $r - 1$ variables
ϵ_j = the error associated with the measurement of control for the jth month

The statistical problems related to the application of this technique to a set of data are concerned with the estimation of the B's, with tests that permit conclusions as to whether a particular B_i is worth keeping in the linear model and with the determination of how well the model fits the observations. The statistical estimation of the B's is accomplished by the method of least squares.* The relative importance of the selected independent variables can be judged by the numerical values of certain test statistics, viz, F and t. The F-statistic is used to test the hypothesis that all the B's are zero against the alternative that at least one of the B's is different from zero. On the other hand the t-statistic is used to test the importance of a particular B_i.

Another quantity that can be used to summarize the results of a multiple

*This approach chooses those values of the B's that minimized the sum of squares of the deviations between the recorded V_j and those computed using the linear model.

regression analysis is the statistic R^2. This quantity measures how much of the observed variability in the recorded proportion of the population subject to VC control is explained by the linear model. Although the numerical value of R^2 can assume values between 0 and 1, values of less than 0.8 are considered unreliable in the analysis. A computed value of R^2 near 1 means that the model is explaining most of the variability in the reported VC control.

The study went through two distinct phases: The first attempt at modeling the conflict was restricted by a limited amount of data at the province level (i.e., 12 months) and by a concept requiring the simultaneous incorporation of a large number of variables. Specifically, an effort was made to "combine" the data from all the provinces and months into a single regression analysis. The standard multiple regression method based on the province-month data produced results that, although informative, were of unacceptable quality, observably unreliable, and subject to numerous pitfalls. In search of an improved model a method developed by Duncan and Walser was introduced.[2] The calculations of control improved by a factor of 10, but even these results remained unsatisfactory. The details of the foregoing investigations are described in App B. These results are included for two reasons: (a) they show in some detail why the methods did not work acceptably and why any future investigations probably should not follow these paths, and (b) at the same time, given situations where the paucity of data may be commensurate with that with which the analysis was initiated, the second method (Duncan and Walser) may well prove useful.

The second phase of the investigation was an analysis on a separate province-by-province basis (as against combined). It began concurrently with the acquisition of additional data; i.e., for purposes of this report, 22 months of data in all relevant categories. The results of these separate province analyses constitute the principal part of this report since these proved to be the most useful.

Regression Analysis for Long An Province

Although results are available in detail for all 43 provinces, they are summarized here in a highly condensed form. A more detailed analysis is included for a single province selected at random (frontispiece, Key 25, Long An) so that several important aspects of a multiple regression analysis can be illustrated. The variable to be measured is VC Controlled. The available data consist of the reported VC Controlled population for 22 months and a corresponding set of independent variables shown in Table 6. For Long An the following four models are discussed:

Model 1: dependent variable, VC Controlled; independent variables: date, large operations, small operations, ARVN, RF, PF, VC, and VC military incidents

Model 2: dependent variables, VC Controlled; independent variables: ARVN, RF, PF, and VC forces

Model 3: dependent variables, VC Controlled, weighted; independent variables, same as Model 1

Model 4: dependent variable, VC Controlled, weighted; independent variables, same as Model 2

For Long An Province the numerical values of the F-statistic are summarized in Table 7. When the computed F-values are compared with their corresponding critical values the results are found to be significant in the statistical sense at the 0.001 level. Thus the likelihood that an event has occurred is either less than 1 time in 1000 or at least one of the B's is different from zero. This statement is true for all four models. In effect the results show that VC control is influenced by or is related to at least one of the independent variables being considered. However, these F-tests fail to identify the particular independent variables that are worth keeping in the linear model. To gain insight as to which of these regressors are important, the t-statistic is utilized as summarized in Table 8. Numerical values of t greater than 3 are significant at the 0.01 level.

An examination of these computed t-values implies that ability to measure VC control in Long An Province is probably due to the relation between control and the number of RF and PF present. In the case of Model 4 (weighted, with four regressors), as the RF forces increased by 1000 men, VC control declined by 4.6 percentage points; whereas, when the number of PF forces increased by 1000 men, VC control declined by 2.6 percentage points. The analysis indicated, but does not "prove," a negative relation between VC control and RF and PF forces. The results show that a regression analysis can suggest important relations, identify historical patterns, and provide clues that will ultimately lead to an understanding of cause-and-effect relations; but statistical "proof," if it exists, can only be accomplished under a controlled experimental situation.

The results also reveal how well the multiple regression model fits the observations. The quantity R^2 measures the extent to which the model explains the observed variability in control; for the four models under consideration the numerical values of R^2 were 0.92, 0.88, 0.98, and 0.93, respectively. Since values greater than 0.80 are considered to be reliable these results imply that all four regression models explain much of the variability observed in VC control. When the dependent variable is VC Controlled, weighted, the model appears to fit even better.

During the 22-month study period the average value of VC Controlled (unweighted) in Long An Province was 0.4073. The standard error of the estimate of this value was 0.0063, giving a relative error for an average computed value of VC Controlled of about 2 percent. Table 9 contains a detailed summary of the month-by-month deviations between the reported and computed values of VC Controlled. The errors obtained when the "combined" and Duncan-Walser models were used are also included to illustrate the relatively poor quality of these approaches.

An examination of the errors shows the superiority of the single-province approach. Not only are the errors smaller, but they exhibit a more random distribution. The Duncan-Walser analysis displays smaller differences than the combined analysis, but the plus errors occur at the beginning of the time period whereas the negative errors occur at the end. In App B some explanation of this lack of randomness is presented. Basically this result appears due to the heterogeneity of the regressor coefficients from one province to another.

It is clear that for Long An Province the analytical model could be used

TABLE 7

Analysis of Variance for Long An Province[a]

Source	Sum of squares	Degrees of freedom	Mean square	F-test
Eight Independent Variables, Y is VC Controlled				
Regression	0.005860	8	0.0007325	17.9137
Error	0.0005316	13	0.0000409	—
Total	0.006391	21	0.0003044	—
Four Independent Variables, Y is VC Controlled				
Regression	0.005606	4	0.001401	30.3136
Error	0.000786	17	0.000046	—
Total	0.006391	21	0.000304	—
Eight Independent Variables, Y is VC Controlled, Weighted				
Regression	0.02334	8	0.002918	99.0226
Error	0.000383	13	0.000029	—
Total	0.02373	21	0.001130	—
Four Independent Variables, Y is VC Controlled, Weighted				
Regression	0.002217	4	0.005543	60.6802
Error	0.001553	17	0.000091	—
Total	0.02373	21	0.001130	—

[a]Key number 25.

TABLE 8

Summary of Regression Analysis for Long An Province[a]

Independent variable	VC Controlled: Partial regression coefficients $\hat{\beta} \times 10^3$	VC Controlled: Ratio to standard error $\frac{\hat{\beta}}{SE(\hat{\beta})}$	VC Controlled, weighted: $\hat{\beta} \times 10^3$	VC Controlled, weighted: $\frac{\hat{\beta}}{SE(\hat{\beta})}$
Eight Variables				
Date	−0.229	−0.305	−3.058	−4.789
Large operations	−0.630	−1.435	−0.508	−1.364
Small operations	0.226	0.573	−0.021	−0.062
ARVN	0.002	0.496	0.001	0.416
RF	−0.014	−1.357	−0.007	−0.786
PF	−0.017	−2.788	−0.025	−4.754
VC	0.033	1.982	0.005	0.337
VC military incidents	0.188	1.766	0.230	2.539
Four Variables				
ARVN	−0.003	−1.041	−0.005	−1.425
RF	−0.027	−7.389	−0.046	−8.758
PF	−0.021	−4.014	−0.026	−3.544
VC	0.029	1.819	−0.011	−0.514

[a]Key number 25.

for the period studied in place of the in-progress standard field method of measuring control.

(U) Can the model be used for prediction over future time periods? At this stage in the investigation the model can be used to help anticipate the effect of changes, for example, in the force structure within a province, especially as long as conditions do not change to any great degree. However, predicting the

(C) TABLE 9

Comparison of VC Controlled Residuals

(Province 25, Long An)

Month	Observed Y	Standard deviation	Duncan's model deviations	Province by province[a]
1964				
May	0.439	-0.00872	0.026	0.00079
Jun	0.439	-0.06502	0.012	0.00429
Jul	0.432	-0.03211	0.032	0.00512
Aug	0.425	0.07137	-0.004	-0.00290
Sep	0.418	0.11267	-0.001	-0.00707
Oct	0.415	0.04427	-0.006	-0.00538
Nov	0.417	-0.01048	-0.005	0.00503
Dec	0.412	-0.08992	-0.028	-0.00037
1965				
Jan	0.408	0.03492	-0.005	0.00148
Feb	0.408	0.08177	-0.004	-0.00046
Mar	0.409	-0.05483	-0.015	-0.00263
Apr	0.406	0.02688	-0.002	0.00610
May	0.400	—	—	-0.00094
Jun	0.400	—	—	-0.00177
Jul	0.392	—	—	-0.00221
Aug	0.387	—	—	0.00447
Sep	0.383	—	—	0.00271
Oct	0.378	—	—	-0.01396
Nov	0.383	—	—	-0.00585
Dec	0.393	—	—	0.00466
1966				
Jan	0.402	—	—	0.00460
Feb	0.414	—	—	0.00429

[a]Nine variables.

level of VC control by the use of a multiple regression model should be considered at least hazardous because the model may not contain all the relevant independent variables. For example, the VC may decide to withdraw its troops from a province independent of any allied activity in the region. The model does not contain a variable measuring VC objectives.

(U) The main results are highlighted for all the provinces in App C. A summary of the results of all these analyses for all the provinces is presented in Tables 10 and 11. From Table 10 it is clear that over 85 percent of the regressions fitted on a province-by-province basis had F-values that were significant in the statistical sense. These results suggest that for the large majority of the provinces a multiple regression model could have been used to approximate VC control, with the provinces taken separately, for both VC

Controlled, unweighted and weighted. It should be noted that over 10 percent of the provinces exhibited little or no change in VC control during the study period. For these provinces lack of statistically significant regression fittings could have been anticipated.

TABLE 10

Summary of *F*-Values for 43 Separate Province Regressions

F-value	VC Controlled		VC Controlled, weighted	
	With date	5 major	With date	5 major
$F < F_{0.001}$	9	8	5	6
$F \geq F_{0.001}$	34	35	38	37

The problem of deciding which regressor coefficients and subsequently which independent variables are worth keeping in the linear model is more complex, especially when 43 provinces are being considered simultaneously. Different independent variables appear as important regressors in different provinces. Table 11 provides some insight into this problem. In this table the t-values for the 43 separate regressions are summarized. Those independent variables with the largest number of statistically significant t-values are the ones most likely to be useful regressors. Thus date, ARVN, RF, PF, and VC forces stand out as important independent variables. Only the variable

TABLE 11

Summary of *t*-Values for 43 Separate Province Regressions[a]

Variable factor	VC Controlled		VC Controlled, weighted	
	With date	5 major	With date	5 major
Date	13 (6)	—	19 (13)	—
Large operations	0 (0)	—	3 (0)	—
Small operations	1 (0)	—	1 (0)	—
ARVN	6 (3)	11 (8)	4 (0)	12 (7)
RF	2 (1)	16 (11)	5 (3)	21 (15)
PF	7 (1)	11 (6)	12 (7)	11 (7)
CIDG	3 (3)	7 (3)	4 (3)	5 (5)
VC	4 (1)	10 (7)	6 (4)	13 (8)
VC military incidents	3 (1)	—	4 (0)	—

[a] $t \geq t_{0.01}$; values in parenthses are for $t_{0.001}$.

date is difficult to explain; it acts as a catchall variable. An examination of the unadjusted pairwise correlations of date with the other independent variables reveals that it is highly correlated with several other important regressors, viz, large operations and RF. When date was removed as one of the regressor variables the results remained essentially unchanged. Some variables like VC military incidents did not stand out as important based on their respective t-values, but care should be exercised in dropping any variable from future regression analyses.

RESULTS

The results can best be appreciated when viewed in the context of a basic research investigation. In contrast to the amount of research effort expended on conventional and nuclear warfare, only a meager effort has been applied to the kind of conflict going on in Vietnam. There are at best only a few benchmarks against which to assess the results of this or kindred studies. Finally the usefulness of the study and the results can best be judged only when possible specific applications are precisely defined. The results judged to be most interesting and relevant are presented in following paragraphs.

The first objective of the study, to develop and test a mathematical model of the conflict in Vietnam, is judged to have been achieved. The statistical test results described in the preceding section, in particular Tables 7 to 9, and in App C, were satisfactory. The principal negative results are also discussed in App B. Because of the heterogeneity of the areas of interest (the provinces), they cannot be combined when employing the standard linear multiple regression method; a more elaborate multiple regression method (Duncan-Walser approach) does provide better results; and the standard method appears to be adequate when the provinces are evaluated separately.

As a part of the overall analysis the vagaries of the field-reported data were examined. Appendix A includes the details of the data adjustments made to overcome apparent or evident periodic inconsistencies.

Population and area control was chosen as the principal criterion of pacification progress. Pertinent data were reported in five separate categories and, as a part of the overall analysis, a method was developed to combine these into a single quantitative value that reflected the qualitative definitions taken together (see App A).

It was concluded that the field-generated data were sufficiently consistent (after adjustments and some gap filling) to be used as a basis for evaluating the computed data from the analytical model. The results are summarized in Tables 3 and 4, and are exhibited in detail in App D. The apparent precision achieved by the model led to further conclusions.

With a possible exception of a few provinces the model could have been used as an alternative method to the field observations made routinely in Vietnam. Examples supporting this result include the detailed examination of Long An Province in the text, the examination of both Long An and Binh Thuan Provinces in App A, and the overview of results, including countrywide and corps results in App D. Data are available (in voluminous computer printouts) covering all 43 provinces in Vietnam.

The results of the model's computations were sufficiently precise that the trends in population control, i.e., the direction and rate of change, could be observed reliably. Such trend observations are, of course, of special interest when the forecasting capabilities of the model are considered.

The factors (independent variables) that best relate to and affect population control were identified (see Table 11). This result, of course, is limited to the types and number of variables selected and used. The manipulation of the model in association with the application of logic can identify the factors that best measure the dependent variable and also discover root causes.

Effects of changes in the values of the independent variables (e.g., the

number of RVN RF and PF) on population control can be calculated. The potential for evaluating force capabilities and requirements is obvious. Should the study be continued, it would naturally explore this capability more extensively.

The overall analysis, including the process of adjusting the basic data to remove inconsistencies as described in App A, indicates that the field-reporting systems could be monitored advantageously for the early detection and correction of data discrepancies.

The results indicate that linear multiple regression analysis is a useful quantitative method for measuring pacification progress in Vietnam (as defined in this study), and that the overall analytical model could be used advantageously to check and improve the various ongoing data-reporting systems. This study could be extended, employing the current model, to make it current with events in Vietnam.

The evidence exhibited by the results, including the degree of precision achieved, appears to support the view that a more comprehensive model could be developed to measure pacification progress even more precisely, to identify and quantify the causal relations, to estimate force capabilities and requirements, and for prediction purposes.

Appendix A

EXAMPLES OF DATA ADJUSTMENT AND EMPLOYMENT OF WEIGHTING FACTORS

REPRESENTATIVE PROVINCES

As noted previously, the provinces of Vietnam differ in area; type of terrain; productivity; population size, distribution, and density; social and religious structure; ethnic bases; political viability and orientation; and friendly and enemy intentions during selected time periods. No single province can be considered typical; however, Long An Province and Binh Thuan Province are studied in detail here as representative areas.

Tables A1 to A4 present detailed data for these provinces as examples of the use of the data-adjustment method and the employment of the weighting factors. Should the need arise, basic data have been preserved for the possible construction of approximately 120 additional tables for the other provinces, special zones, division and corps tactical zones, VC military regions, and for the entire country.

A principal and interesting characteristic was that Long An had been under VC domination for over 20 years, and efforts (over 32 months) to eradicate VC influence had only very modest success, overall about 8.5 percent (see Table A2). This was so even though its proximity to Saigon made it important to the GVN over the entire period. It tended to reflect the general situation in the delta. As for Binh Thuan, on the other hand, no more than one-third of the province was initially credited to the VC; and over the period, especially during 1966, VC overall control (VC Controlled, weighted and adjusted) decreased by almost a factor of 3 (see Table A4). The reporting system in both provinces took 3 to 6 months to begin to reflect the real (in historical perspective) situation, and nearly a year to reach a state where only minor adjustments appeared to be appropriate. (This was also typical countrywide, by province.)

(C)

TABLE A1

Reported and Adjusted Population Control Data for Long An Province

(May 1964–December 1966)

Date	Population base, thous	Secured: Reported population Thousands	Secured: Reported population Percent	Secured: Adjusted population Thousands	Secured: Adjusted population Percent	Undergoing Securing: Reported population Thousands	Undergoing Securing: Reported population Percent	Undergoing Securing: Adjusted population Thousands	Undergoing Securing: Adjusted population Percent	Undergoing Clearing: Reported population Thousands	Undergoing Clearing: Reported population Percent	Undergoing Clearing: Adjusted population Thousands	Undergoing Clearing: Adjusted population Percent	Uncontested: Reported population Thousands	Uncontested: Reported population Percent	Uncontested: Adjusted population Thousands	Uncontested: Adjusted population Percent	VC Controlled: Reported population Thousands	VC Controlled: Reported population Percent	VC Controlled: Adjusted population Thousands	VC Controlled: Adjusted population Percent
1964																					
May	382.9	0.0	0.0	6.8	1.7	84.2	22.0	71.6	18.8	26.8	7.0	136.4	35.6	3.8	1.0	0.0	0.0	268.1	70.0	168.1	43.9
Jun	382.9	0.0	0.0	13.2	3.4	84.2	22.0	65.2	17.1	26.8	7.0	136.4	35.6	3.8	1.0	0.0	0.0	268.1	70.0	168.1	43.9
Jul	382.9	0.0	0.0	19.6	5.1	40.5	10.6	61.4	16.1	57.3	14.9	136.4	35.6	3.8	1.0	0.0	0.0	281.3	73.5	165.5	43.9
Aug	383.0	28.2	7.4	26.0	6.8	24.7	6.4	57.7	15.1	49.9	13.0	136.4	35.6	3.8	1.0	0.0	0.0	276.4	72.2	162.9	42.5
Sep	383.0	28.2	7.4	32.4	8.4	24.7	6.4	53.9	14.1	49.9	13.0	136.4	35.6	3.8	1.0	0.0	0.0	276.4	72.2	160.3	41.9
Oct	379.1	28.2	7.6	38.8	10.2	72.8	19.5	46.2	12.2	113.6	30.4	136.4	36.0	3.8	1.0	0.0	0.0	155.1	41.5	157.7	41.6
Nov	378.5	33.4	9.0	45.2	11.9	67.3	18.1	39.2	10.3	113.6	30.4	136.4	36.1	3.8	1.0	0.0	0.0	155.1	41.5	157.7	41.7
Dec	378.5	53.2	14.2	51.6	13.6	25.8	6.9	34.6	9.1	136.4	36.5	136.4	36.1	3.8	1.0	0.0	0.0	154.3	41.4	155.9	41.2
1965																					
Jan	378.5	58.0	15.5	58.0	15.3	23.1	6.2	31.9	8.4	134.2	35.9	134.2	35.5	3.8	1.0	0.0	0.0	154.4	41.4	154.4	40.8
Feb	378.5	58.0	15.5	64.6	17.1	24.7	6.7	27.0	7.1	132.6	35.5	132.6	35.0	3.8	1.0	0.0	0.0	154.3	40.8	154.3	40.8
Mar	378.5	58.0	15.4	71.2	18.8	24.7	6.5	20.0	5.3	132.6	35.0	132.6	35.0	3.8	1.0	0.0	0.0	159.3	42.1	154.7	40.9
Apr	378.5	60.7	16.3	77.8	20.5	27.7	7.4	14.2	3.8	127.8	34.2	133.0	35.1	3.8	1.0	0.0	0.0	153.5	40.6	153.5	40.6
May	379.6	84.6	22.3	84.6	22.3	7.2	1.8	9.7	2.6	142.2	37.5	133.4	35.1	3.8	1.0	0.0	0.0	141.8	37.3	151.9	40.0
Jun	379.6	84.6	22.3	84.6	22.3	7.2	1.8	10.9	2.9	142.2	37.5	133.8	35.2	3.8	1.0	0.0	0.0	141.8	37.3	150.3	39.6
Jul	379.6	85.2	22.4	85.2	22.4	19.2	2.4	11.4	3.0	129.6	34.1	134.3	35.4	3.8	1.0	0.0	0.0	141.8	37.3	148.7	39.2
Aug	379.6	85.2	22.4	85.2	22.4	13.9	3.7	12.6	3.3	134.8	35.5	134.8	35.0	3.8	1.0	0.0	0.0	141.8	37.3	146.9	38.7
Sep	379.5	85.2	22.4	85.2	22.5	13.9	3.7	14.2	3.7	134.8	35.5	134.8	35.5	3.8	1.0	0.0	0.0	141.8	37.3	145.3	38.3
Oct	379.5	85.6	22.6	85.6	22.6	10.8	2.8	12.8	3.4	137.4	36.2	137.4	36.2	3.8	1.0	0.0	0.0	141.8	37.3	143.6	37.8
Nov	379.4	85.6	22.6	85.6	22.6	10.8	2.8	11.0	2.9	137.4	36.2	137.4	36.2	3.8	1.0	0.0	0.0	141.8	37.3	145.4	38.3
Dec	374.6	87.8	23.4	87.8	23.4	10.0	2.7	4.1	1.1	135.6	36.2	135.6	36.2	0.0	0.0	0.0	0.0	141.2	37.7	147.1	39.3
1966																					
Jan	379.4	87.8	23.1	87.8	23.1	10.0	2.6	3.6	0.9	135.6	35.7	135.6	35.8	4.8	1.2	0.0	0.0	141.2	37.2	152.4	40.2
Feb	380.5	81.1	21.3	86.2	22.6	13.2	3.5	5.1	1.3	140.2	36.8	131.5	34.6	4.8	1.2	0.0	0.0	141.2	37.2	157.7	41.5
Mar	382.6	81.1	21.2	84.6	22.1	6.2	1.6	7.5	2.0	127.4	33.3	127.4	33.2	4.8	1.2	0.0	0.0	163.1	42.6	163.1	42.7
Apr	380.6	83.0	21.8	83.0	21.8	4.3	1.1	9.1	2.4	125.4	32.9	125.4	32.9	4.8	1.2	0.0	0.0	163.1	42.8	163.1	42.9
May	380.6	83.3	21.9	83.3	21.9	5.0	1.3	9.8	2.5	124.4	32.7	124.4	32.7	4.8	1.2	0.0	0.0	163.1	42.8	163.1	42.9
Jun	379.5	85.6	22.6	85.6	22.6	1.7	0.4	6.5	1.7	124.6	32.8	124.6	32.8	4.8	1.2	0.0	0.0	162.8	42.9	162.8	42.9
Jul	379.5	85.6	22.6	85.6	22.6	2.3	0.6	7.1	1.9	123.7	32.6	123.7	32.6	4.8	1.2	0.0	0.0	163.1	42.9	163.1	42.9
Aug	380.6	87.7	23.0	87.7	23.0	5.2	1.4	5.2	1.4	119.8	31.5	124.6	32.7	4.8	1.2	0.0	0.0	163.1	42.9	163.1	42.9
Sep	380.9	89.3	23.4	89.3	23.4	3.9	1.0	8.7	2.3	119.8	31.5	119.8	31.4	4.8	1.2	0.0	0.0	163.1	42.9	163.1	42.9
Oct	379.1	90.2	23.9	90.2	23.8	5.4	1.4	5.4	1.4	115.2	30.4	120.0	31.6	4.8	1.2	0.0	0.0	163.1	43.0	163.1	43.0
Nov	378.7	91.0	24.0	91.0	24.0	4.9	1.3	4.9	1.3	115.2	30.4	120.0	31.7	4.8	1.2	0.0	0.0	162.8	43.0	162.8	43.0
Dec	378.7	91.0	24.0	91.0	24.0	4.9	1.3	4.9	1.3	115.2	30.4	120.0	31.7	4.8	1.2	0.0	0.0	162.8	43.0	162.8	43.0

(C)

TABLE A2

Percentages of Reported and Adjusted Population Control Data and Calculated Residuals for Long An Province

(May 1964–December 1966)

Date	VC Controlled			VC Controlled, weighted			RVN Controll weighted	
	Reported	Adjusted	Calculated residuals[a]	Reported	Adjusted	Calculated residuals[a]	Reported	Adju
1964								
May	70.0	43.9	−0.14	81.8	70.9	0.76	18.2	29
Jun	70.0	43.9	0.94	81.8	70.4	0.74	18.2	29
Jul	73.5	43.9	0.47	86.6	69.4	0.55	13.4	30
Aug	72.2	42.5	−0.16	82.9	68.4	0.62	17.1	31
Sep	72.2	41.9	−0.88	82.9	67.4	−0.50	17.1	32
Oct	41.5	41.6	−0.67	65.6	66.8	−0.67	34.4	33
Nov	41.5	41.7	0.54	65.3	66.4	0.23	34.7	33
Dec	41.4	41.2	0.12	65.4	65.6	−0.92	34.6	34
1965								
Jan	41.4	40.8	0.01	64.9	64.6	0.03	35.1	35
Feb	40.8	40.8	−0.49	64.5	63.9	−0.45	35.5	36
Mar	42.1	40.9	−0.04	66.0	63.5	−0.71	34.0	36
Apr	40.6	40.6	0.64	64.0	62.8	0.72	36.0	37
May	37.3	40.0	−0.20	61.4	61.9	−0.60	38.6	38
Jun	37.3	39.6	−0.03	61.4	61.6	0.20	38.6	38
Jul	37.3	39.2	0.02	60.3	61.3	0.03	39.7	38
Aug	37.3	38.7	0.27	60.8	61.0	1.33	39.2	39
Sep	37.3	38.3	0.13	60.8	60.7	1.10	39.2	39
Oct	37.3	37.8	−1.28	61.0	60.6	−0.60	39.0	39
Nov	37.3	38.3	−0.64	61.0	60.9	−1.21	39.0	39
Dec	37.7	39.3	0.50	60.2	61.3	0.40	39.8	38
1966								
Jan	37.2	40.2	0.39	60.7	61.9	−0.63	39.3	38
Feb	37.2	41.5	0.48	61.5	62.6	−0.42	38.5	37
Mar	42.6	42.7	—	64.3	63.3	—	35.7	36
Apr	42.8	42.9	—	64.2	63.3	—	35.8	36
May	42.8	42.9	—	64.1	63.2	—	35.9	36
Jun	42.9	42.9	—	64.0	63.1	—	36.0	36
Jul	42.9	42.9	—	64.0	63.1	—	36.0	36
Aug	42.9	42.9	—	63.4	62.9	—	37.1	37
Sep	42.9	42.9	—	63.2	62.4	—	37.6	37
Oct	43.0	43.0	—	62.9	62.4	—	37.6	37
Nov	43.0	43.0	—	62.9	62.4	—	37.6	37
Dec	43.0	43.0	—	62.9	62.4	—	37.6	37

[a]The differences between the adjusted control percentages and those obtained by the multiple reg sion analysis, with eight variables, 22-month test.

(C)

TABLE A3

Reported and Adjusted Population Control Data for Binh Thuan Province

(May 1964–December 1966)

Date	Population base, thous	Secured				Undergoing Securing				Undergoing Clearing				Uncontested				VC Controlled			
		Reported population		Adjusted population		Reported population		Adjusted population		Reported population		Adjusted population		Reported population		Adjusted population		Reported population		Adjusted population	
		Thousands	Percent	Thousands	Percent	Thousands	Percent	Thousands	Percent	Thousands	Percent	Thousands	Percent	Thousands	Percent	Thousands	Percent	Thousands	Percent	Thousands	Percent
1964																					
May	244.5	8.1	3.3	90.6	36.6	192.0	78.5	50.0	20.2	41.0	16.7	98.5	39.7	0.0	0.0	0.0	0.0	3.4	1.4	8.6	3.5
Jun	244.5	8.1	3.3	90.6.	36.6	192.0	78.5	50.0	20.2	41.0	16.5	97.5	39.3	0.0	0.0	0.0	0.0	3.4	1.4	9.7	3.9
Jul	247.7	57.8	23.3	90.6	36.6	145.6	58.8	50.0	20.2	40.9	16.5	96.3	38.8	0.0	0.0	0.0	0.0	3.4	1.4	10.8	4.4
Aug	247.7	71.1	28.7	90.6	36.6	131.9	53.2	50.0	20.2	40.9	16.5	95.2	38.4	0.0	0.0	0.0	0.0	3.8	1.5	11.9	4.8
Sep	247.7	90.6	36.6	98.6	39.8	112.3	45.3	42.0	16.9	41.0	16.5	94.1	38.0	0.0	0.0	0.0	0.0	3.8	1.5	13.0	5.3
Oct	247.7	101.8	41.1	106.6	43.1	49.1	19.8	34.0	13.7	93.0	37.5	93.0	37.5	0.0	0.0	0.0	0.0	3.8	1.5	14.1	5.7
Nov	247.7	122.2	49.3	114.6	46.3	28.7	11.6	24.9	10.1	93.0	37.5	93.0	37.5	0.0	0.0	0.0	0.0	3.8	1.5	15.2	6.1
Dec	247.7	130.6	52.7	122.6	49.5	23.9	9.6	16.3	6.6	84.5	34.1	92.5	37.3	4.8	1.9	0.0	0.0	3.8	1.5	16.3	6.6
1965																					
Jan	247.7	130.6	52.7	130.6	52.7	23.9	9.6	8.2	3.3	84.5	34.1	91.5	36.9	4.8	1.9	0.0	0.0	3.8	1.5	17.4	7.1
Feb	247.7	154.1	62.2	131.0	52.9	7.7	3.1	7.7	3.1	82.0	33.1	90.5	36.5	0.0	0.0	0.0	0.0	3.8	1.5	18.5	7.5
Mar	247.7	118.4	47.8	135.0	54.5	0.0	0.0	2.4	1.0	123.1	49.7	90.7	36.6	2.4	1.0	0.0	0.0	3.8	1.5	19.6	7.9
Apr	247.7	118.4	47.8	137.2	55.4	0.0	0.0	0.0	0.0	123.1	49.7	89.8	36.2	2.4	1.0	0.0	0.0	3.8	1.5	20.7	8.4
May	248.5	119.2	48.0	139.4	55.8	0.0	0.0	0.0	0.0	118.1	47.5	88.7	35.5	7.4	2.9	0.0	0.0	3.8	1.5	21.8	8.7
Jun	252.7	141.4	55.9	141.4	55.9	0.0	0.0	2.8	1.1	87.9	34.8	84.9	33.6	0.0	0.0	0.0	0.0	23.6	9.3	23.6	9.4
Jul	252.7	141.4	55.9	141.4	55.9	0.0	0.0	6.6	2.6	87.9	34.8	81.1	32.1	0.0	0.0	0.0	0.0	23.6	9.3	23.6	9.4
Aug	252.7	141.4	55.9	141.4	55.9	0.0	0.0	10.4	4.1	87.9	34.8	77.3	30.6	0.0	0.0	0.0	0.0	23.6	9.3	23.6	9.4
Sep	252.8	141.4	55.9	141.4	55.9	1.3	0.5	14.4	5.7	86.6	34.2	73.5	29.1	0.0	0.0	0.0	0.0	23.5	9.3	23.5	9.3
Oct	252.8	144.5	57.1	144.5	57.1	15.1	6.0	15.1	6.0	69.7	27.6	69.7	27.6	0.0	0.0	0.0	0.0	23.6	9.3	23.6	9.3
Nov	252.9	144.5	57.1	144.5	57.1	15.1	6.0	15.1	6.0	69.7	27.6	69.7	27.6	0.0	0.0	0.0	0.0	23.5	9.3	23.5	9.3
Dec	252.8	144.5	57.1	144.5	57.1	15.1	6.0	15.1	6.0	69.7	27.6	69.7	27.6	0.0	0.0	0.0	0.0	23.5	9.3	23.5	9.3
1966																					
Jan	252.8	145.9	57.7	145.9	57.7	1.7	0.7	23.3	9.2	82.1	32.5	60.5	23.9	0.0	0.0	0.0	0.0	23.2	9.1	23.2	9.2
Feb	252.9	151.8	60.0	151.8	60.0	3.6	1.4	27.1	10.7	51.5	20.4	51.5	20.4	0.0	0.0	0.0	0.0	45.9	18.2	22.4	8.9
Mar	252.8	157.2	62.2	157.2	62.2	7.5	2.9	21.6	8.5	56.2	22.3	52.4	20.7	0.0	0.0	0.0	0.0	31.9	12.6	21.6	8.6
Apr	252.8	163.6	64.7	163.6	64.7	30.9	12.2	15.1	6.0	53.3	21.1	53.3	21.1	0.0	0.0	0.0	0.0	5.0	2.0	20.8	8.2
May	252.8	164.7	65.1	164.7	65.1	29.8	11.8	20.8	8.2	53.3	21.1	47.3	18.7	0.0	0.0	0.0	0.0	5.0	2.0	20.0	8.0
Jun	253.6	173.8	68.5	173.8	68.5	21.5	8.5	19.3	7.6	53.3	21.1	41.3	16.3	0.0	0.0	0.0	0.0	5.0	2.0	19.2	7.6
Jul	251.7	173.8	69.0	173.8	69.0	24.4	9.7	24.4	9.7	35.4	14.1	35.4	14.1	0.0	0.0	0.0	0.0	18.1	7.2	18.1	7.2
Aug	251.7	173.8	69.0	182.7	72.6	24.4	9.7	20.1	8.0	35.4	14.1	35.4	14.1	0.0	0.0	0.0	0.0	18.1	7.2	13.5	5.3
Sep	251.7	173.8	69.0	191.6	76.1	24.4	9.7	15.8	6.3	35.4	14.1	35.4	14.1	0.0	0.0	0.0	0.0	18.1	7.2	8.9	3.5
Oct	261.3	200.6	76.7	200.6	76.7	11.4	4.4	11.4	4.4	49.3	18.8	49.3	18.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nov	261.3	200.6	76.7	200.6	76.7	11.4	4.4	11.4	4.4	42.9	16.4	49.3	18.9	6.4	2.4	0.0	0.0	0.0	0.0	0.0	0.0
Dec	261.3	200.6	76.7	200.6	76.7	11.4	4.4	11.4	4.4	49.3	18.9	49.3	18.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

(C)

TABLE A4

Percentages of Reported and Adjusted Population Control Data and Calculated Residuals for Binh Thuan Province

(May 1964–December 1966)

Date	VC Controlled			VC Controlled, weighted			RVN Controlled, weighted	
	Reported	Adjusted	Calculated residuals[a]	Reported	Adjusted	Calculated residuals[a]	Reported	Adjusted
1964								
May	1.4	3.5	-0.22	35.0	33.4	-0.41	65.0	66.6
Jun	1.4	3.9	-0.17	35.0	33.6	-0.09	65.0	66.4
Jul	1.4	4.4	-0.25	32.9	33.8	0.28	67.1	66.2
Aug	1.5	4.8	0.13	27.4	33.9	0.25	72.6	66.1
Sep	1.5	5.3	-0.09	25.1	33.2	0.71	74.9	66.8
Oct	1.5	5.7	0.11	30.0	32.3	0.12	70.0	67.7
Nov	1.5	6.1	0.03	27.5	31.7	0.36	72.5	68.3
Dec	1.5	6.6	0.11	26.8	30.9	-0.75	73.2	69.1
1965								
Jan	1.5	7.1	-0.46	26.8	30.2	0.16	73.2	69.8
Feb	1.5	7.5	0.01	22.3	30.3	0.20	77.7	69.7
Mar	1.5	7.9	0.61	32.3	30.2	-1.11	67.7	69.8
Apr	1.5	8.4	0.19	32.3	30.1	-0.24	67.7	69.9
May	1.5	8.7	0.12	33.0	30.0	0.24	67.0	70.0
Jun	9.3	9.4	0.68	30.2	29.8	0.42	69.8	71.2
Jul	9.3	9.4	-0.33	30.2	29.4	1.04	69.8	70.6
Aug	9.3	9.4	-0.76	30.2	28.9	0.16	69.8	71.1
Sep	9.3	9.3	0.12	30.0	28.4	-1.22	70.0	71.6
Oct	9.3	9.3	-0.25	27.7	27.7	0.14	72.3	72.3
Nov	9.3	9.3	-0.00	27.6	27.6	0.19	72.4	72.4
Dec	9.3	9.3	0.49	27.6	27.6	0.64	72.4	72.4
1966								
Jan	9.1	9.2	0.22	28.8	26.3	0.08	71.2	73.7
Feb	18.2	8.9	-0.30	30.8	24.3	-1.17	69.2	75.7
Mar	12.6	8.6	—	26.8	23.6	—	73.2	76.4
Apr	2.0	8.2	—	18.3	22.6	—	81.7	77.4
May	2.0	8.0	—	18.2	21.7	—	81.8	78.3
Jun	2.0	7.6	—	17.1	19.7	—	82.9	80.3
Jul	7.2	7.2	—	18.5	18.6	—	81.5	81.4
Aug	7.2	5.3	—	18.5	16.2	—	81.5	83.8
Sep	7.2	3.5	—	18.5	13.9	—	81.5	86.1
Oct	0.0	0.0	—	12.6	12.7	—	87.4	87.3
Nov	0.0	0.0	—	13.6	12.7	—	86.4	87.3
Dec	0.0	0.0	—	12.6	12.7	—	87.4	87.3

[a]The differences between the adjusted control percentages and those obtained by the multiple regression analysis, with eight variables, 22-month test.

Appendix B

MULTIPLE REGRESSION MODELS

EARLY MODELS

This appendix includes a relatively detailed discussion of the multiple regression models used at the beginning of the investigation, partly because of a paucity of data. They have been judged to be somewhat inadequate for the reasons given, and exposure of the inadequacies may be expected to alert others to the limitations discovered.

The use of a multiple regression model is based on the supposition that a variable such as the proportion of the population subject to VC Control* can be measured more precisely when it is expressed as a function of several apparently related variables such as population density, VC activity, and the number of friendly forces. A multiple regression model is usually based on a linear model. Here it is meant linear in the parameters, where a parameter refers to any unknown quantity of the model that can vary over a set of values. Several distinct models are introduced here.

The first two models represent an attempt to analyze the factors related to VC Control on a countrywide basis by province, using the entire set of available data. Model 1 represents a standard multiple regression approach where the parameters are the partial regression coefficients. They measure the change in the dependent variable (say, proportion under VC Control) that is due to a specific independent variable after adjusting for the effects of the remaining variables under consideration. The independent variables are those variables believed to relate to or affect VC Control.

Model 2 is an extension of the first model to the case where there is assumed to be an unknown province effect. This method was developed by D. B. Duncan and M. Walser.[2] It provides a procedure for simultaneously estimating the regression parameters and the variance components for the within- and between-province errors. Furthermore, analysis of the data suggested that Model 2 did not completely account for the observed province effects, and an alternative approach was required. A third method (or model) that appears useful and adequate is to use the standard multiple regression approach with the data for the provinces, taking them one at a time.

Model 1: A Standard Multiple Regression Model

Suppose that an observation is made on a specified geographical area (a province within South Vietnam) at a fixed point in time. The objective of this observation is to measure the proportion of the population in this province subject to VC Control. Although possibly subject to an unknown bias, a procedure

*Except as specifically indicated, the expression "VC Control" as used throughout this section means unweighted VC population and area control as reported.

was developed that would permit quantitative measurement of this value. For convenience, this observation was denoted by the letter y. Further, it was assumed that there was knowledge of the numerical values of several other related factors, such as population density and the number of friendly forces in the province. These known factors were denoted by the letters $x_1, \ldots, x_r$, where r is the number of independent variables under consideration. Finally it was noted that the measurement of control is subject to an error, denoted by the letter ϵ.

This model assumes that the dependent variable y can be expressed as an additive function (or linear combination) of the independent variables and ϵ. Thus the model can be written

$$y_\ell = \beta_o + \beta_1 x_{1\ell} + \beta_2 x_{2\ell} + \ldots + \beta_r x_{r\ell} + \epsilon_\ell \tag{1}$$

where β_o = the underlying level of control

β_i = the change in y due to the ith independent variable $x_{i\ell}$, after adjusting for the remaining $r - 1$ variables $i = 1, \ldots, r$

ϵ_ℓ = the error associated with the ℓth observation on y

$\ell = 1, \ldots, n$

The statistical estimation of the parameters of this model is based on the principle of least squares. The least-squares method is a procedure by which the parameters are estimated by minimizing a certain quadratic form of the observations. The best estimators of the partial regression coefficients are obtained by minimizing the sum of squares of the deviations of the observed data from the fitted model.

Now consider the case where there are n observations on control (either for different provinces or the same province for several months, or a combination of both). The model can be restated in more formal mathematical terms. The matrix notation, $y' = (y_1, \ldots, y_n)$ is a vector of observations and a matrix of independent variables X. The model can be written

$$y = X\beta + \epsilon \tag{2}$$

where y and ϵ are $n \cdot 1$ column vectors,

X = a $n \cdot k$ matrix of known coefficients, and

β = a $k \cdot 1$ column vector of unknown partial regression coefficients.

The basic assumptions are:

(1) The elements of the X matrix are known constants.

(2) The independent variables are uncorrelated.

(3) The expected value of ϵ is 0.

(4) The variance of ϵ is $I\sigma^2$ where I is an $n \cdot n$ identity matrix and σ^2 is the common error of measurement. The errors are also assumed to be uncorrelated and independent of the elements of X.

The method of least squares shows that if $n \geq k$ and $|x'x| \neq 0$ the minimum variance unbiased estimator of β is given by

$$\hat{\beta} = (X'X)^{-1} X'y \tag{3}$$

and the variance covariance matrix for $\hat{\beta}$ is

$$\mathrm{var}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$$

where σ^2 is estimated from the residual sum of squares divided by the total number of observations n minus the number of partial regression coefficients r minus 1. That is,

$$\hat{\sigma}^2 = (y - X\hat{\beta})(y - X\hat{\beta})/(n - r - 1) \tag{4}$$

If the assumptions of the model are justified, the hypothesis that $\beta_1 = \beta_2 = \ldots = \beta_r = 0$ can be tested against the alternative that at least one $\beta_i \neq 0$ using the test statistic

$$F = \hat{\beta}'X'y/r\hat{\sigma}^2 \tag{5}$$

where F has an F distribution with r and $n - r - 1$ degrees of freedom.

This model is based on the assumption that the independent variables are known constants and that they are not subject to any appreciable errors of measurement. When the independent variables are population density, time period (month), the number of RVN operations, etc, this assumption is reasonable. When the independent variable involves VC activities and VC forces this assumption may not be strictly satisfied, e.g., owing to errors in intelligence or reporting. For the purpose of this analysis it is assumed that the possible errors in the independent variables are small relative to the variation observed in y.

The assumption that the errors are homogeneous is not strictly valid for another reason. Recall that the dependent variable is the proportion of the population subject to VC Control. The error has the form $p(1 - p)/N$ where p is the proportion controlled in a given province at a fixed point in time and N is the corresponding total population size. Only if the true proportion of control did not differ greatly and the population sizes were about the same from province to province could the errors be completely homogeneous. This, of course, is not the case. The usual methods of weighted regression and the use of an inverse sine transformation when applied to the available data in conjunction with a standard regression analysis apparently do not have an appreciable homogenizing effect on the errors; consequently these approaches are not discussed in this report.

Model 2: The Duncan-Walser Approach[2,3]

An observation is now denoted by y_{ij} for the jth observation in the ith province. The model can be written

$$y_{ij} = \alpha + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + \beta_r x_{rij} + \alpha_i + \epsilon_{ij} \tag{6}$$

$$i = 1, \ldots, p$$

$$j = 1, \ldots, T$$

where $x_{\ell ij}$ = the ℓth regressor for β_ℓ a fixed unknown partial regression coefficient $\ell = 1, \ldots, r$

α = a fixed unknown intercept

α_i = the uncorrelated random province (plot) errors with common variance σ_α^2 $(i = 1, \ldots, p)$

ϵ_{ij} = the uncorrelated (subplot) errors with common variance σ_ϵ^2

Associated with this model is a parameter $\rho = \sigma_\alpha^2/\sigma_\epsilon^2$ a ratio of the between-province variation to the within-province variation.

If ρ were known, and the errors normally distributed, the quadratic form for the sum of the "within," the "between," and the "mean" components would be

$$S_w^2/\sigma^2 = \sum_{i=1}^{p}\sum_{j=1}^{T}(\tilde{y}_{ij} - \beta_1\tilde{x}_{1ij} - \ldots - \beta_r\tilde{x}_{rij})^2/\sigma_\epsilon^2 \tag{7}$$

$$S_B^2/\sigma_\epsilon^2 = \sum_i W_i(\tilde{y}_i - \beta_i\tilde{x}_{1i} - \ldots - \beta_r\tilde{x}_{ri})^2/\sigma_\epsilon^2 \tag{8}$$

$$S_X^2/\sigma_\epsilon^2 = \sum_i W_i(\bar{y} - \mu)^2/\sigma_\epsilon^2 \tag{9}$$

The curl terms are the deviations

$$\tilde{y}_{ij} = y_{ij} - \bar{y}_i\,,\ \tilde{x}_{kij} = x_{kij} - \bar{x}_{ki}\,,\ \tilde{y}_i = \bar{y}_i - \bar{y}\,,\ \tilde{x}_{ki} = x_{ki} - \bar{x}_k$$

and from the plot (province) means

$$\bar{y}_i = \sum_j y_{ij}/n_i\,,\ \bar{x}_{ki} = \sum_j x_{kij}/n_i$$

$$\bar{y} = \sum_i W_i y_i/\Sigma W_i\,,\ \bar{x}_k = \sum_i W_i x_{ki}/\Sigma W_i$$

where

$$W_i = (\rho + 1/n_i)^{-1}$$

and the overall weighted mean

$$\mu = E(\bar{y}) = \alpha + \beta_1\bar{x}_1 + \beta_2\bar{x}_2 + \ldots + \beta_r\bar{x}_r \tag{10}$$

The maximum likelihood estimates of the regression coefficients can be obtained from the combined normal equations given by

$$A_C b_C = g_C \tag{11}$$

where A_C and g_C are the combined $r \cdot r$ and $r \cdot 1$ matrices of weighted sums of deviation products. That is

$$A_C = A_W + A_B \text{ and } g_C = g_W + g_B \tag{12}$$

where

$$a_{hk}^W = \sum_{i=1}^{p}\sum_{j=1}^{T}\tilde{x}_{hij}\tilde{x}_{kij}\,,\ a_{hk}^B = \sum_{i=1}^{p}W_i\tilde{x}_{hi}\tilde{x}_{ki} \tag{13}$$

$$g_k^W = \sum_{i=1}^{p}\sum_{j=1}^{T}\tilde{x}_{kij}\tilde{y}_{ij} \text{ and } g_k^B = \sum_{i=1}^{p}W_i\tilde{x}_{ki}\tilde{y}_i \tag{14}$$

The maximum likelihood estimate of μ is y and that of α is

$$a = \bar{y} - b_1\bar{x}_1 - \ldots - b_r\bar{x}_r \tag{15}$$

The detailed theoretical developments related to Model 2 are given in Duncan and Walser's paper and will not be reproduced here. Note that this model requires the estimation of the variance component ratio ρ in order to estimate the weights W_i . These weights are required for the between analysis and the combined analysis.

Results

In this section the two multiple regression models described above are applied to a select set of Vietnamese data compiled on a province-month basis. The complete set of data consisted of 516 vectors of information of the form $(x_{1 2}, x_{2\ell}, \ldots, x_{r\ell}, y_\ell)$ corresponding to 43 provinces and 12 consecutive months from May 1964 to April 1965. A summary of the data is presented in Table B1.

TABLE B1

Summary of Data[a]

(May 1964–April 1965)

Variable factor	Mean	Standard deviation	Minimum	Maximum
d_k	−0.287	0.602	−1.021	0.893
Date	6.500	3.455	1.000	12.000
Large operations	4.824	5.067	0.000	33.000
Small operations	7.814	8.953	0.000	104.000
ARVN	1643.112	1430.808	0.000	7,277.000
RF	1864.607	666.117	463.000	3,883.000
PF	3672.919	2149.292	525.000	10,580.000
CIDG	487.225	583.331	0.000	2,685.000
VC forces	899.089	979.478	0.000	4,620.000
VC military incidents	26.516	24.786	0.000	141.000
VC control	0.212	0.172	0.000	0.642
Weighted VC control	0.435	0.178	0.029	0.751

[a]Of the (current) 44 provinces in South Vietnam, 43 were included in the analysis. [In December 1966, Sa Dec was formed by including portions of Vinh Binh (No. 36, frontispiece) and Vinh Long (No. 37, frontispiece)]

The first 10 variables represent the independent variables: month (scaled 1 through 12), population density (an index that compares each province to an average density for the country as a whole), the numbers of friendly forces (subdivided into ARVN, RF, PF, and CIDG), VC activity, and estimated number of VC forces. The final element of the vector is the numerical value of VC Control.

When Model 1 was employed, the data were treated as if each of the 516 province-month vectors represented independent observations. On the other hand, when Model 2 was used, the data were treated in a manner similar to a split-plot experiment where each province represented a "whole plot" and the months within provinces corresponded to "subplots." The superiority of the latter approach is demonstrated in the following analysis.

For comparative purposes the estimated partial regression coefficients $\hat{\beta}$ and the ratio of these estimates to their standard errors $|\hat{\beta}|/SE(\hat{\beta})$ are given

as well as the corresponding analysis-of-variance tables. In addition an analysis of the computed residual errors $y - \hat{y}$ is used to dramatize the difficulties inherent in each model and as a means of identifying those basic assumptions that appear unjustified.

Tables B2 to B4 summarize the analysis associated with unweighted VC Control as the dependent variable and the standard multiple regression model. Tables B5 and B6 correspond to an application of the Duncan-Walser Model on VC Control. A comparison of the residual errors for a typical province (Dinh Tuong, 29, frontispiece) is presented in Table B7. The remaining tables represent a second analysis using weighted VC Control as the dependent variable (App A).

From the analysis of variance given in Table B2, note that at least one partial regression coefficient is different from zero. If the assumptions are valid the chance of observing a calculated F 38.87 with 10 and 505 degrees of freedom is less than 1 in 1000 when all $\hat{\beta}$'s are zero. The estimates of the $\hat{\beta}$'s and the ratio of these estimates to their standard errors $\left[|\hat{\beta}|/SE(\hat{\beta})\right]$ are given in Table B3. A ratio with a numerical value greater than two is taken to mean that this particular $\hat{\beta}$ is significantly related to VC Control. The value of $\hat{\beta}_\ell$ is the change in VC Control per unit of x_ℓ after adjusting for all other independent variables under investigation.

The data support the hypothesis that positive relations exist between VC Control and large (friendly) operations, VC forces, and VC military activities. On the other hand, negative relations were noted between VC Control and small (friendly) operations, and numbers of RF and PF. The corresponding regression coefficients are significant in a statistical sense only if the underlying assumptions are valid.

In Table B4 the partial regression coefficients are presented in a standardized form that makes each independent variable independent of the particular range of the data and treats each variable as if the range were 0 to 1. After adjusting the data to this standard form, note that both VC military incidents and large operations have a strong positive relation to VC Control. The effect of the other regressors appears to be less than one-half the effect of either of these two variables. The fact that VC Control is going up in regions where VC military incidents have increased is not surprising. But for VC Control to be increasing where RVN large operations are increasing is not obvious. One explanation might be that the number of large operations has not increased sufficiently to counteract the increasing VC Control. There exists the possibility of a quadratic relation between the number of RVN forces and VC Control. The current analysis does not take this fact into account; however, future attempts to structure the conflict should.

Table B5 gives the steps and results associated with Model 2, the Duncan-Walser model. The procedure is sequential and preliminary steps are required to estimate the weights W_i used in the between-regression analysis and to test for homogeneity of the between- and within-regression analysis. The final combined analysis yields a computed F of 23.09 with 10 and 505 degrees of freedom. A statistically significant result is obtained at the 0.001 level assuming the underlying assumptions are valid. Consequently it is concluded that at least one $\hat{\beta}_C$ is not zero.

TABLE B2

Summary of Standard Regression Analysis for VC Control: Analysis of Variance

Source of variation	Degree of freedom	Sums of squares	Mean squares	F
Regression	10	6.627	0.663	38.873
Error	505	8.609	0.017	—
Total	515	15.236	—	—

TABLE B3

Summary of Standard Regression Analysis on Provinces for VC Control: Ratio of Partial Regression Coefficients to Their Standard Error

Independent variable	Partial regression coefficients $\hat{\beta} \times 10^3$	Ratio to standard error $\frac{\lvert\hat{\beta}\rvert}{SE(\hat{\beta})}$
d_k	14.400	0.89
Date	0.754	0.44
Large operations	11.610	7.80
Small operations	−3.244	1.10
ARVN	0.007	1.10
RF	−0.029	2.51
PF	−0.009	2.53
CIDG	0.008	0.67
VC forces	0.019	2.62
VC military incidents	2.676	8.39

TABLE B4

Summary of Standard Regression Analysis for VC Control: Partial Regression Coefficients in Standardized Form

Independent variable	Standard form, $\hat{\beta} \times 10^2$
d_k	5.040
Date	1.515
Large operations	34.202
Small operations	−16.886
ARVN	5.537
RF	−11.395
PF	−11.735
CIDG	2.662
VC forces	10.795
VC military incidents	38.557

TABLE B5

Between- and Within-Regression Analysis for VC Control: Steps and Results

Weight	Within analysis	Between analysis 1	Between analysis 2	Combined analysis	Final analysis: Between	Final analysis: Combined
Regression degree of freedom	9	10	10	10	10	10
Error degree of freedom	464	32	32	505	32	505
EMS, ($\times 10^3$)	1.415	34.180	1.415	1.494	1.344	1.485
F-ratio	26.692	3.938	3.938	22.978	3.938	23.098

TABLE B6

Between- and Within-Regression Analysis for VC Control: Summary of Combined Analysis

Independent variable	Partial regression coefficients $\hat{\beta} \times 10^3$	Ratio to standard error $\frac{\lvert\hat{\beta}\rvert}{SE(\hat{\beta})}$
d_k	98.695	2.721
Date	3.328	5.236
Large operations	3.097	5.044
Small operations	0.411	1.594
ARVN	0.008	1.935
RF	-0.036	4.772
PF	-0.026	5.154
CIDG	-0.032	3.246
VC forces	-0.006	1.257
VC military incidents	0.165	1.242

TABLE B7

Comparison of Residual Errors for VC Control, Province 29

Month	Observed y	Standard	Duncan
1964			
May	0.540	0.152	0.036
Jun	0.540	0.218	0.034
Jul	0.535	-0.059	0.011
Aug	0.535	0.101	0.032
Sep	0.511	0.004	0.011
Oct	0.490	-0.002	-0.003
Nov	0.468	0.069	0.003
Dec	0.447	-0.015	-0.026
1965			
Jan	0.425	-0.015	-0.029
Feb	0.425	0.173	-0.019
Mar	0.425	0.073	-0.028
Apr	0.420	0.122	-0.021

The estimated partial regression coefficients used in the combined analysis and the associated ratios are given in Table B6. These estimates contrast sharply with those obtained under Model 1 and an application of the standard multiple regression technique. More specifically, the population density deviation (which now depends only on the between-province analysis) is found to be positively related to VC Control and an average increase in VC Control over the 12 months under investigation is also indicated. The effect of large RVN operations and VC military incidents are much less significant. However, the effects of RF, PF, and CIDG continue to exhibit a strong negative relation with VC Control.

TABLE B8

Summary of Standard Regression Analysis for Weighted VC Control: Analysis of Variance

Source of variation	Degree of freedom	Sums of squares	Mean squares	F
Regression	10	7.457	0.746	42.500
Error	505	8.861	0.018	—
Total	515	16.317	—	—

TABLE B9

Summary of Standard Regression Analysis for Weighted VC Control: Ratio of Partial Regression Coefficients to Their Standard Error

Independent variable	Partial regression coefficients $\hat{\beta} \times 10^3$	Ratio to standard error $\frac{\lvert\hat{\beta}\rvert}{SE(\hat{\beta})}$
d_k	-2.732	0.17
Date	-6.537	3.73
Large operations	10.054	6.64
Small operations	-3.499	4.19
ARVN	0.031	5.12
RF	-0.046	3.91
PF	0.008	2.01
CIDG	0.066	5.53
VC forces	0.024	3.28
VC military incidents	2.185	6.74

Tables B8 to B12 keep the independent variables as they were in the preceding analyses but use a weighted VC Control as the dependent variable. The weights attempt to account for the fact that regions identified as Undergoing Securing or Undergoing Clearing also are subject to varying degrees of VC Control. The numerical values of the weights were 0.3, 0.6, and 1.0, respectively, for Undergoing Securing, Undergoing Clearing, and VC Controlled. Although their weights are arbitrary they incorporate the best judgment available on the meaning of each category of control.

The analysis of variance for weighted VC Control based on Model 1 is given in Table B8. The computed F was 42.50, a statistically significant re-

TABLE B10

Summary of Standard Regression Analysis for Weighted VC Control: Partial Regression Coefficients in Standardized Form

Independent variable	Standard form, $\hat{\beta} \times 10^2$
d_k	−0.924
Date	−12.688
Large operations	28.620
Small operations	−17.599
ARVN	25.276
RF	−17.409
PF	9.143
CIDG	21.566
VC forces	13.303
VC military incidents	30.431

TABLE B11

Between- and Within-Regression Analysis for Weighted VC Control: Steps and Results

Weight	Within analysis	Between analysis		Combined analysis	Final analysis	
		1	2		Between	Combined
Regression degree of freedom	9	10	10	10	10	10
Error degree of freedom	464	32	32	505	32	505
EMS ($\times 10^3$)	1.440	35.705	1.440	1.540	1.375	1.531
F ratio	12.234	4.280	4.280	10.155	4.280	10.203

TABLE B12

Between- and Within-Regression Analysis for Weighted VC Control: Summary of Combined Analysis

Independent variable	Partial regression coefficients $\hat{\beta} \times 10^3$	Ratio to standard error $\frac{\lvert\hat{\beta}\rvert}{SE(\hat{\beta})}$
d_k	64.833	1.743
Date	−2.713	4.203
Large operations	1.939	3.110
Small operations	0.409	1.560
ARVN	0.012	2.850
RF	−0.008	1.025
PF	−0.020	3.773
CIDG	−0.031	3.096
VC forces	0.000	0.057
VC military incidents	0.379	2.809

sult indicating that the partial regression coefficients are not all zero. The estimates of the $\hat{\beta}$'s and the associated ratios are given in Table B9. As in the unweighted case, a positive relation appears to exist between control and large operations. In contrast to the unweighted case the time effect appears negative, indicating a decline in weighted control over the time period under investigation.

REASONS FOR INADEQUACY

The question can now be asked, on what basis should the two models previously discussed be judged? An analysis of the deviations observed between the data and the fitted equation can provide an answer. The residual errors indicated that neither model is entirely adequate. To illustrate this point a detailed look is taken at the errors associated with the application of each model to the data. Table B7 contains the computed errors for Province 29 (a typical case). During the period May 1964–April 1965, Province 29 experienced a decline in VC Control starting at 0.54 and ending at 0.42.

When Model 1 was employed the errors ranged from a value of -0.0591 to a value of +0.2180. On the other hand, when Model 2 was introduced the errors ranged from -0.0291 to +0.0356. An examination of the residuals associated with the remaining 42 provinces gave similar results. In general the errors associated with Model 1 were greater than those computed using Model 2.

At first glance one might conclude that Model 2 was an improved method of measuring VC Control and, indeed, to some extent this conclusion is justified. However, a further examination of these errors suggests that the residuals associated with Model 2 tend to be nonrandom. There is a strong tendency for the errors to be clumped with the plus (+) values at the beginning of the observational period and with the negative (-) values at the end. In some provinces the residuals appear positively correlated, whereas in others they appear negatively correlated. In either case this represents a violation of an underlying assumption, viz, that the errors are random.

One explanation for this appears to be that both models are based on the hypothesis that regression behaves the same way from province to province. One obvious violation of this hypothesis is the regressor "date." As date increases from 1 to 12 the change in VC Control may increase or decrease. The result of applying the current models to the data is to estimate a regressor coefficient that is <u>not</u> representative of the relations being studied. Since similar phenomena are plausible with each regressor (independent) variable being considered, extreme caution is necessary when analyzing data on a country-wide basis. The pooled regression may be inappropriate for describing the relations of interest from one province to the next.

Unless a great deal more is learned about the relations between the dependent and independent variables in a conflict of the type being studied, one should not attempt to pool the data to produce a single regressor model for the entire country. Instead, separate regressions should be applied to the data collected at the province level (or smaller geographical grouping). A combined or pooled analysis might result as a second step after provinces with homogeneous regressions have been identified. However, each pooled analysis should be followed by a detailed statement that justifies the reasons for combining the provinces into a single regression equation.

Appendix C

STANDARD MULTIPLE REGRESSION ANALYSES OF VC CONTROL

SIGNIFICANCE OF DATA

The purpose of this appendix is to provide additional information on the kinds of data analyzed in the study, especially the reported percentages of VC population and area control for the period covered at the province, special zone, division, and corps tactical levels of operation.

These data are included because they underline the following significant considerations: (a) the quality of heterogeneity of the separate areas, with respect to control, that must be accommodated by the analytical model; (b) the ranges of changes in control to be analyzed; (c) the need to study the problem in detail, at least down to the province level, rather than to depend on aggregated, countrywide or special zone, division, and corps levels if the most useful results are to be obtained; (d) the differences in the picture of the conflict, according to the echelon observed; and (e) simplicity of the calculations required to combine the province data in order to derive the results for whatever aggregation levels may be required.

Province Level

The four models previously discussed for Long An Province are presented here in highly condensed form for each province. In Tables C1 to C4 an xx implies that the corresponding statistic, F or t, exceeded the critical value for the 0.001 level of significance, thus being considered very highly significant in the statistical sense for the total regression or the particular regression coefficient. The sign, plus or minus, is included where appropriate to indicate the direction of the regression coefficient. An x indicates significance at a level between 0.001 and 0.01. The computed t- and F- statistics that did not show significance even at the 0.01 level are not marked.

Above Province Level

The values in Tables C5 and C6 were computed by simple addition of the values for the included provinces.

For the 22-month period May 1964–February 1966, the percentage changes for the two categories of VC Controlled, unweighted and weighted, and for the various echelons are listed in Table C7. These data indicate the ranges over which the model must operate. Of course, the significance of the changes will depend also on the range of control levels at the beginning and end of the period. As expected, the ranges of percentage changes are observed to diminish on balance as the data are aggregated. The picture of the conflict also changes

(C)

TABLE C1

Summary of Statistical Results of Province-by-Province Standard Multiple Regression Analyses: I Corps

Key number[a]	Province	Division	Approximate average population,[b] thous	Percent reported VC Controlled or weighted VC Controlled: May 64	Percent reported VC Controlled or weighted VC Controlled: Feb 66	VC variable measured		F^c	R^d	Date	Large operations	Small operations	ARVN	RF	PF	CIDG	VC main and local forces	VC military incidents
1	Quang Tri	1	270	1.9	2.3	Unweighted	All	—	0.72	—	—	—	—	—	—	—	—	—
							5 major	×	0.62	—	—	—	—	—	—	×	—	—
				47.5	36.6	Weighted	All	××	0.86	—	×	—	—	—	−××	—	—	—
							5 major	××	0.72	—	—	—	—	−×	−××	—	—	—
2	Thua Thien	1	460	0.2	4.7	Unweighted	All	××	1.00	××	—	—	−×	—	−××	—	−×	—
							5 major	××	0.85	—	—	—	−×	—	−××	−×	—	—
				29.0	33.4	Weighted	All	××	0.99	××	—	—	—	−××	−××	—	—	—
							5 major	××	0.79	—	—	—	—	—	−××	−×	—	—
3	Quang Nam	2	570	4.6	61.0	Unweighted	All	××	0.99	—	—	—	—	—	—	—	—	—
							5 major	××	0.99	—	—	—	—	−××	−×	−×	—	—
				50.3	74.4	Weighted	All	××	0.99	—	—	—	—	−×	−××	−××	—	—
							5 major	××	0.99	—	—	—	—	−××	−××	−××	—	—
4	Quang Tin	2	355	12.5	48.8	Unweighted	All	××	0.98	×	—	—	—	—	—	—	—	—
							5 major	××	0.95	—	—	—	—	—	−×	—	—	—
				63.8	68.8	Weighted	All	××	0.93	—	—	—	—	—	—	—	—	—
							5 major	××	0.88	—	—	—	—	—	—	—	—	—
5	Quang Ngai	2	650	32.5	45.1	Unweighted	All	××	0.93	—	—	—	—	—	—	—	—	—
							5 major	××	0.92	—	—	—	×	×	—	—	—	—
				51.7	52.3	Weighted	All	—	0.70	—	—	—	—	—	—	—	—	—
							5 major	—	0.58	—	—	—	—	—	—	—	—	—

[a]As on frontispiece.
[b]Rural population only, 13,075,400; six cities (1,811,600, rounded) not included.
[c]× indicates the test statistic exceeds the 0.01 critical value F or t; ×× corresponds to the 0.001 level.
[d]Observed variability.

CONFIDENTIAL

(C)

TABLE C2

Summary of Statistical Results of Province-by-Province Standard Multiple Regression Analyses: II Corps

Key number[a]	Province	Division	Approximate average population,[b] thous	Percent reported VC Controlled or weighted VC Controlled: May 64	Percent reported VC Controlled or weighted VC Controlled: Feb 66	VC variable measured		F^c	R^d	Date	Large operations	Small operations	ARVN	RF	PF	CIDG	VC main and local forces	VC military incidents
6	Binh Dinh	22	850	27.9	38.7	Unweighted	All	xx	0.91	—	—	—	−x	—	—	—	—	—
							5 major	xx	0.86	—	—	—	—	x	—	xx	—	—
				62.7	63.9	Weighted	All	—	0.68	—	—	—	—	—	—	—	—	—
							5 major	x	0.60	—	—	—	—	—	—	xx	—	—
7	Phu Bon	22	45	20.3	21.8	Unweighted	All	—	0.60	—	—	—	—	—	—	—	—	—
							5 major	—	0.46	—	—	—	—	—	—	—	—	—
				34.7	48.8	Weighted	All	x	0.84	—	—	—	—	—	—	—	—	—
							5 major	xx	0.81	—	—	—	x	—	—	—	—	—
8	Pleiku	24SZ	160	17.9	54.0	Unweighted	All	xx	0.97	x	—	—	—	−x	—	—	—	—
							5 major	xx	0.91	—	—	—	—	−xx	—	—	—	—
				41.5	66.3	Weighted	All	xx	0.96	x	—	—	—	—	—	—	—	—
							5 major	xx	0.90	—	—	—	—	−xx	—	—	—	—
9	Kontum	24SZ	105	18.2	29.9	Unweighted	All	xx	0.96	—	—	—	—	—	—	—	−xx	—
							5 major	xx	0.87	—	—	—	—	−x	—	—	—	—
				38.9	47.1	Weighted	All	xx	0.99	—	—	—	—	—	—	xx	xx	−x
							5 major	xx	0.96	—	—	—	—	−x	—	xx	xx	—
10	Quang Duc	23	30	7.6	7.5	Unweighted	All	xx	0.82	—	—	—	—	—	—	—	—	—
							5 major	xx	0.78	—	—	—	—	—	−xx	—	—	—
				26.8	26.8	Weighted	All	xx	0.92	—	—	—	—	—	−x	—	—	—
							5 major	xx	0.87	—	—	—	—	—	−xx	xx	—	—
11	Darlac	23	200	3.6	9.9	Unweighted	All	xx	0.97	xx	—	—	−x	—	—	—	—	—
							5 major	xx	0.87	—	—	—	−x	—	—	—	—	—
				50.4	58.8	Weighted	All	xx	0.96	xx	—	—	—	—	—	—	—	—
							5 major	xx	0.73	—	—	—	—	—	—	—	—	—
12	Phu Yen	22	340	14.1	33.4	Unweighted	All	xx	0.95	—	—	—	—	—	—	—	—	—
							5 major	xx	0.93	—	—	—	—	−xx	—	—	xx	—
				63.2	58.9	Weighted	All	x	0.78	—	—	—	—	—	—	—	—	—
							5 major	x	0.70	—	—	—	—	−xx	—	—	—	—

CONFIDENTIAL

13	Khanh Hoa	23	295	2.7	2.5	Unweighted	All	×	0.78	—	—	—	—	—	—	—	—	—
							5 major	××	0.75	—	—	—	—	—	×	—	—	—
				30.1	17.8	Weighted	All	××	0.98	−×	—	—	—	—	−×	—	—	—
							5 major	××	0.95	—	—	—	—	−××	—	−ˇ	−××	—
14	Tuyen Duc	23	80	1.5	1.7	Unweighted	All	—	0.58	—	—	—	—	—	—	—	—	—
							5 major	—	0.38	—	—	—	—	—	—	—	—	—
				17.1	9.9	Weighted	All	××	0.91	—	—	—	—	—	—	—	—	—
							5 major	××	0.83	—	—	—	×	—	—	—	—	—
15	Ninh Thuan	23	150	1.8	1.8	Unweighted	All	—	0.00	—	—	—	—	—	—	N[e]	—	—
							5 major	—	0.00	—	—	—	—	—	—	N[e]	—	—
				22.7	10.8	Weighted	All	××	0.87	−××	—	—	—	—	−××	N[e]	—	—
							5 major	—	0.41	—	—	—	—	—	—	N[e]	—	—
16	Lam Dong	23	65	3.2	9.5	Unweighted	All	××	0.95	—	—	—	—	—	−×	N[e]	—	—
							5 major	××	0.93	—	—	—	—	—	−××	N[e]	—	—
				18.3	25.7	Weighted	All	××	0.88	—	—	—	—	—	—	N[e]	—	—
							5 major	××	0.83	—	—	—	—	—	−×	N[e]	—	—
17	Binh Thuan	23	250	3.5	8.9	Unweighted	All	××	0.98	×	—	—	—	—	—	N[e]	—	—
							5 major	××	0.96	—	—	—	−××	××	−××	N[e]	××	—
				33.4	24.3	Weighted	All	××	0.97	—	—	—	—	—	—	N[e]	—	—
							5 major	××	0.93	—	—	—	×	−××	××	N[e]	−××	—

[a]As on frontispiece.
[b]Rural population only, 13,075,400; six cities (1,811,600, rounded) not included.
[c]× indicates the test statistic exceeds the 0.01 critical value F or *t*; ×× corresponds to the 0.001 level.
[d]Observed variability.
[e]No CIDG reported in province.

(C)

TABLE C3

Summary of Statistical Results of Province-by-Province Standard Multiple Regression Analyses: III Corps[a]

Key number[b]	Province	Division	Approximate average population,[c] thous	Percent reported VC Controlled or weighted VC Controlled: May 64	Percent reported VC Controlled or weighted VC Controlled: Feb 66	VC variable measured		F^d	R^e	Date	Large operations	Small operations	ARVN	RF	PF	CIDG	VC main and local forces	VC military incidents
18	Binh Tuy	10	65	0.6	12.4	Unweighted	All	xx	0.97	—	—	—	—	—	−x	—	—	—
							5 major	xx	0.95	—	—	—	−xx	—	—	—	—	—
				30.0	29.4	Weighted	All	xx	0.92	—	—	—	—	—	—	—	—	—
							5 major	xx	0.86	—	—	—	−xx	—	—	—	—	—
19	Binh Long	5	80	19.8	25.6	Unweighted	All	xx	0.95	xx	—	—	—	—	—	—	—	—
							5 major	xx	0.81	—	—	—	—	—	—	—	—	—
				54.5	41.4	Weighted	All	xx	0.93	−x	—	—	—	—	—	—	—	—
							5 major	xx	0.82	—	—	—	—	—	—	—	—	—
20	Phuoc Long	5	65	2.2	58.1	Unweighted	All	xx	0.99	x	—	—	—	—	−x	−xx	—	x
							5 major	xx	0.93	—	—	—	—	xx	—	—	x	—
				38.3	63.7	Weighted	All	xx	0.97	—	—	—	—	—	—	−xx	—	x
							5 major	xx	0.92	—	—	—	—	xx	—	−xx	—	—
22	Long Khanh	10	125	0.0	6.0	Unweighted	All	xx	0.93	—	—	—	—	—	—	N[f]	—	—
							5 major	xx	0.76	—	—	—	−x	xx	—	N[f]	—	—
				15.2	18.1	Weighted	All	xx	0.91	—	—	—	—	—	—	N[f]	—	—
							5 major	xx	0.79	—	—	—	−xx	xx	—	N[f]	—	—
23	Bien Hoa	10	310	0.2	2.1	Unweighted	All	xx	0.99	—	—	−x	xx	—	—	—	—	—
							5 major	xx	0.98	—	—	—	xx	—	—	—	x	—
				32.4	13.4	Weighted	All	xx	0.98	−x	—	—	—	—	—	—	—	—
							5 major	xx	0.94	—	—	—	−xx	—	—	—	−x	—
24	Phuoc Tuy	10	100	13.3	17.8	Unweighted	All	xx	0.93	—	—	—	—	—	—	N[f]	—	—
							5 major	xx	0.91	—	—	—	—	xx	—	N[f]	xx	—
				36.2	36.0	Weighted	All	xx	0.84	—	—	—	—	−x	—	N[f]	—	—
							5 major	xx	0.83	—	—	—	x	−xx	—	N[f]	−xx	—
25	Long An	25	380	43.9	39.2	Unweighted	All	xx	0.92	—	—	—	—	—	—	N[f]	—	—
							5 major	xx	0.88	—	—	—	—	−xx	−xx	N[f]	—	—
				70.9	62.6	Weighted	All	xx	0.98	−xx	—	—	—	—	−xx	N[f]	—	—
							5 major	xx	0.93	—	—	—	—	−xx	−x	N[f]	—	—

26	Hau Nghia	25	225	50.4	48.2	Unweighted	All	××	0.86	—	—	—	—	—	—	—	—	—
							5 major	×	0.63	—	—	—	—	—	×	—	—	—
				74.3	69.9	Weighted	All	××	0.84	−××	—	—	—	—	—	—	—	—
							5 major	×	0.52	—	—	—	—	—	—	—	—	—
27	Tay Ninh	25	235	18.1	10.3	Unweighted	All	××	0.98	—	—	—	—	—	—	−××	—	—
							5 major	××	0.97	—	—	—	××	—	—	−××	—	—
				64.4	23.0	Weighted	All	××	0.99	−××	—	—	—	—	—	—	—	—
							5 major	××	0.95	—	—	—	—	−×	—	—	—	—
28	Binh Duong	5	240	26.9	29.5	Unweighted	All	××	0.86	—	—	—	—	—	—	—	—	—
							5 major	××	0.75	—	—	—	—	−××	—	—	—	—
				56.3	48.2	Weighted	All	××	0.97	−××	—	—	×	—	—	—	—	—
							5 major	××	0.81	—	—	—	—	−×	—	—	−×	—
44	Gia Dinh	CMD	965	0.0	0.0	Unweighted	All	—	0.67	—	—	—	—	—	—	N[f]	−×	—
							5 major	×	0.56	—	—	—	—	—	—	N[f]	−×	—
				12.0	2.0	Weighted	All	××	0.93	−×	—	—	—	—	—	N[f]	—	—
							5 major	××	0.81	—	—	—	—	−××	—	N[f]	×	—

[a]Province 21 deleted; incorporated into four adjacent provinces.
[b]As on frontispiece.
[c]Rural population only, 13,075,400; six cities (1,811,600, rounded) not included.
[d]× indicates the test statistic exceeds the 0.01 critical value F or t; ×× corresponds to the 0.001 level.
[e]Observed variability.
[f]N, no CIDG reported in province.

(C)

TABLE C4

Summary of Statistical Results of Province-by-Province Standard Multiple Regression Analyses: IV Corps

Key number[a]	Province	Division	Approximate average population,[b] thous	Percent reported VC Controlled or weighted VC Controlled: May 64	Percent reported VC Controlled or weighted VC Controlled: Feb 66	VC variable measured		F^c	R^d	Date	Large operations	Small operations	ARVN	RF	PF	CIDG	VC main and local forces	VC military incidents
29	Dinh Tuong	7	520	26.9	29.5	Unweighted	All	xx	0.97	—	—	—	—	—	—	N[e]	—	—
							5 major	xx	0.95	—	—	—	−xx	—	—	N[e]	—	—
				66.0	52.4	Weighted	All	xx	0.97	—	—	—	—	—	—	N[e]	—	—
							5 major	xx	0.97	—	—	—	−xx	—	—	N[e]	−x	—
30	Go Cong	7	175	33.7	21.1	Unweighted	All	xx	0.97	—	—	—	—	—	−x	N[e]	—	—
							5 major	xx	0.96	—	—	—	—	—	−x	N[e]	—	—
				60.7	48.0	Weighted	All	xx	0.99	—	—	—	x	—	−xx	N[e]	—	—
							5 major	xx	0.99	—	—	—	xx	−xx	−xx	N[e]	—	—
31	Kien Hoa	7	540	52.2	42.4	Unweighted	All	xx	0.99	−xx	—	—	—	—	x	N[e]	—	−x
							5 major	xx	0.90	—	—	—	—	−xx	—	N[e]	—	—
				61.8	52.4	Weighted	All	xx	1.00	−xx	−x	—	—	—	x	N[e]	xx	−x
							5 major	xx	0.93	—	—	—	—	−xx	—	N[e]	—	—
32	Kien Tuong	7	60	41.0	37.3	Unweighted	All	—	0.68	—	—	—	—	—	—	—	—	—
							5 major	—	0.57	—	—	—	—	—	—	—	—	—
				47.2	40.5	Weighted	All	xx	0.85	—	—	—	—	—	—	—	—	—
							5 major	xx	0.78	—	—	—	—	—	—	—	—	—
33	An Giang	9	470	9.9	0.0	Unweighted	All	xx	0.99	—	—	—	xx	—	—	—	—	—
							5 major	xx	0.99	—	—	—	xx	—	—	—	−xx	—
				22.0	0.1	Weighted	All	xx	1.00	−xx	—	—	—	—	—	—	−x	—
							5 major	xx	0.99	—	—	—	—	−x	—	—	−xx	—
34	Chau Doc	9	445	5.2	4.0	Unweighted	All	xx	0.94	—	—	—	—	—	—	—	−x	—
							5 major	xx	0.94	—	—	—	—	—	—	x	−xx	—
				31.5	16.2	Weighted	All	xx	0.97	—	—	—	—	−xx	—	—	—	x
							5 major	xx	0.94	—	—	—	x	−xx	—	—	—	—
35	Kien Phong	9	290	25.6	13.5	Unweighted	All	xx	0.96	−x	—	—	—	—	—	—	—	—
							5 major	xx	0.90	—	—	—	—	—	—	—	−xx	—
				48.6	23.0	Weighted	All	xx	1.00	−xx	—	−x	—	—	xx	−xx	—	—
							5 major	xx	0.94	—	—	—	—	−)	—	—	−xx	—

36	Vinh Binh	9	540	19.9	22.3	Unweighted	All	××	0.97	—	—	—	—	××	—	−××	—	—
							5 major	××	0.94	—	—	—	—	×	—	−××	—	—
				55.2	47.7	Weighted	All	××	0.99	−××	—	—	—	—	—	—	—	—
							5 major	××	0.91	—	—	—	—	—	—	—	—	—
37	Vinh Long	9	555	19.5	16.5	Unweighted	All	××	0.88	—	—	—	—	—	—	N[e]	—	—
							5 major	××	0.82	—	—	—	—	−×	—	N[e]	—	—
				30.0	22.6	Weighted	All	××	0.94	—	—	—	—	—	—	N[e]	—	—
							5 major	××	0.92	—	—	—	−××	—	—	N[e]	—	—
38	An Xuyen	21	225	64.2	63.0	Unweighted	All	××	0.93	—	—	—	—	—	—	—	—	—
							5 major	××	0.85	—	—	—	−××	—	—	—	—	—
				71.7	67.5	Weighted	All	××	0.95	−×	—	—	—	—	—	—	—	—
							5 major	××	0.77	—	—	—	—	—	—	×	—	—
39	Ba Xuyen	21	360	28.8	22.7	Unweighted	All	××	0.95	−××	—	—	××	—	—	N[e]	—	—
							5 major	××	0.76	—	—	—	—	−××	—	N[e]	—	—
				52.1	33.7	Weighted	All	××	0.97	−××	—	—	—	—	—	N[e]	—	—
							5 major	××	0.86	—	—	—	—	−××	×	N[e]	—	—
40	Bac Lieu	21	260	34.3	40.9	Unweighted	All	××	0.91	—	—	—	—	—	—	N[e]	—	—
							5 major	××	0.71	—	—	—	—	—	—	N[e]	—	—
				53.1	56.4	Weighted	All	××	0.87	—	—	—	—	×	—	N[e]	−××	—
							5 major	××	0.74	—	—	—	—	××	×	N[e]	−××	—
41	Chuong Thien	21	250	47.0	40.4	Unweighted	All	××	0.99	−××	—	—	—	—	—	—	—	—
							5 major	××	0.94	—	—	—	−××	—	××	—	−××	—
				60.1	56.0	Weighted	All	××	0.98	−××	−×	—	—	—	−×	—	×	—
							5 major	××	0.87	—	—	—	−××	—	—	—	−×	—
42	Kien Giang	9	365	27.8	21.8	Unweighted	All	—	0.73	—	—	—	—	—	—	—	—	—
							5 major	—	0.30	—	—	—	—	—	—	—	—	—
				40.2	34.0	Weighted	All	—	0.74	—	—	—	—	—	—	—	—	—
							5 major	—	0.45	—	—	—	—	—	—	—	—	—
43	Phong Dinh	21	420	40.7	36.1	Unweighted	All	××	0.95	−×	—	—	—	—	×	—	—	—
							5 major	××	0.90	—	—	—	—	—	—	—	—	—
				52.8	44.2	Weighted	All	××	0.97	—	—	—	—	—	×	—	—	—
							5 major	××	0.95	—	—	—	—	—	×	—	−××	—

[a]As on frontispiece.
[b]Rural population only, 13,075,400; six cities (1,811,600, rounded) not included.
[c]× indicates the test statistic exceeds the 0.01 critical value F or t; ×× corresponds to the 0.001 level.
[d]Observed variability.
[e]N, no CIDG reported in province.

(C)

TABLE C5

Reported VC Population and Area Controlled[a]

Corps	Special zones and divisions	Percent VC Controlled		Percentage points change	Percentage change
		May 64	Feb 66		
I		12.4	36.4	24.0	192.4
	1	0.8	3.8	3.0	375.0
	2	17.8	51.5	33.7	189.3
II		14.6	24.6	10.0	69.1
	22	23.8	36.6	12.8	53.8
	24 SZ	18.0	44.4	26.4	146.7
	23	3.0	5.8	2.8	92.7
III	Without CMD[b]	22.9	24.6	1.7	7.4
	With CMD	15.0	16.1	1.1	7.4
	10	2.4	6.6	4.2	175.5
	5	21.2	33.4	12.2	57.5
	25	38.4	33.5	-4.9	-12.8
	CMD	0.0	0.0	0.0	NC
IV		30.7	25.3	-5.4	-17.7
	7	45.8	35.0	-10.8	-23.6
	9	17.3	13.1	-4.2	-24.3
	21	41.3	38.4	-2.9	-6.9
RVN	Without CMD	22.6	27.1	4.5	19.8
	With CMD	21.0	25.1	4.1	19.5

[a]Rural population only; total RVN less the six autonomous cities.
[b]CMD, Capital Military District; mainly Gia Dinh Province.

(C)

TABLE C6

Reported VC Population and Area Controlled, Weighted[a]

Corps	Special zones and divisions	Percent weighted VC Controlled		Percentage points change	Percentage change
		May 64	Feb 66		
I		48.2	54.7	6.5	13.5
	1	35.8	34.5	-1.3	-3.6
	2	53.9	64.0	10.1	18.7
II		47.1	46.7	-0.4	NC
	22	61.8	62.0	0.2	NC
	24 SZ	40.5	58.7	18.2	44.9
	23	31.8	26.1	-5.7	-17.9
III	Without CMD[b]	53.0	41.6	-11.4	-21.5
	With CMD	34.6	27.2	-7.4	-21.5
	10	29.2	19.9	-9.3	-31.9
	5	52.9	49.4	-3.5	-6.6
	25	70.0	53.5	-16.5	-23.6
	CMD	12.0	2.0	-10.0	-83.3
IV		48.7	37.6	-11.1	-22.8
	7	62.6	51.2	-11.4	-18.2
	9	37.4	24.3	-13.1	-35.0
	21	56.8	49.2	-7.6	-13.4
RVN	Without CMD	48.9	37.1	-11.8	-24.1
	With CMD	46.2	34.5	-11.7	-25.3

[a]Rural population only; total RVN less the six autonomous cities under at least nominal RVN Control.
[b]CMD, Capital Military District; mainly Gia Dinh Province.

(C)

TABLE C7

Ranges in Percentage Changes in VC Controlled and VC Controlled, Weighted, for the Various Echelons of Interest

Area	Types of VC Control	Percentage minus	Percentage plus
43 provinces	Controlled	100	2533
	Controlled, weighted	98	66
12 divisions and special zones	Controlled	24	375
	Controlled, weighted	83	44
4 Corps	Controlled	18	192
	Controlled, weighted	23	14
Countrywide	Controlled	0	20
	Controlled, weighted	25	0

somewhat depending on whether one looks at VC Controlled only, or VC Controlled, weighted, which takes into account VC influence in the contested areas (Undergoing Securing and Undergoing Clearing). For example, the VC were increasing their control (i.e., VC Controlled only), especially at the extremes, but their control overall (i.e., VC Controlled, weighted) was decreasing.

(U) VC Controlled, weighted, employed weights of 0.0, 0.3, 0.6, 1.0, as described in the body of the report and App A; RVN Control, weighted, equals 100 - VC Controlled, weighted.

Appendix D

COMPARISON OF ACCURACY OF COMPUTED VALUES
AND REPORTED DATA

SUMMARY OF RESULTS

The establishment of criteria for necessary accuracy and validity of results of the analysis must depend ultimately on the specific problem to be solved and the requirements and judgment of the user or decision maker.

Within the study the results and their acceptability depend in general on a comparison of the official field-reported data (for the rural population) and those independently computed by the multiple regression model. Perspective on the accuracy and validity of the field data has been presented in a number of ways throughout the paper. Although these data are demonstrably imperfect in some detail, as witnessed by the apparent need for some adjustments (see App A), overall pacification progress as measured by field-generated population control data is taken as an acceptable basis for evaluating the tests conducted.

Quantitative Tests

The results and their quality are observed in three kinds of quantitative tests. The first is a determination as to whether the trends over time have meaning and usefulness for possible decision making. In over 60 cases examined, including all provinces and special zones, divisions, corps, and VC military regions countrywide, the trends were traceable. Over 90 percent of the time these trends were essentially unambiguous.* In lieu of presenting all the evidence the detailed data on two provinces are provided in Tables A2 and A4 for Long An and Binh Thuan. In Long An Province there was little change; in Binh Thuan the decrease in VC Controlled (population), weighted, was noteworthy.

The second and third types of results are (a) statistical and (b) direct comparisons of the officially reported data and those computed by the model.

The statistical aspects are presented in detail in the text and in App B, and relate especially to the test statistics F and R^2. The evidence showed that 42 out of the 43 provinces passed the tests; the relatively poor statistical results for one province, Kien Giang (Key No. 42 on frontispiece), located in the Mekong Delta and in the southernmost portion of Vietnam, resisted immediate explanation. The problem of Kien Giang includes possibilities ranging from a possible error in the computer program to simply relatively poor dependent and/or independent variable data.

*Supporting data are available. However, to present all the evidence would require some 60 tables similar to A2, at this time considered to be excessive. In any event, the accuracies attained on a monthly basis as discussed later in this appendix offer implicit proof of trend-analysis reliability.

Accuracy of Computed Values

The balance of this appendix is primarily concerned with the relative accuracy of the computed values as compared with those reported. The results are summarized in Tables D1 to D3, and observations are offered regarding probable reasons for the larger differences encountered.

There are a number of different but closely related ways to express and display results. The choice was to base them on the category VC Controlled, weighted, as this category reflects all the principal steps in the analysis including the application of the multiple regression technique, data-adjustment analysis, and the employment of the weighting method to obtain a single value.

Table D1 is a display of all-province results (for the rural population). The information includes average total population, the ranges in percentages of VC Control during the 22-month test period, May 1964–February 1966, the average VC Control in percent (as well as the number of people), the standard errors of the estimates, and the percentages of relative average differences. Table D2 summarizes the results for all provinces. Table D3 summarizes the differences at countrywide and corps levels of aggregation. Finally a number of itemized probable reasons for the larger differences are provided. Interpretative guidelines follow.

As expected, of course, the differences decrease with an increase in the level of aggregation (Table D3). Countrywide, the relative percentage differences between the computed and reported (after adjustment) results was only about ½ of 1 percent (0.55 percent or 0.60 when Gia Dinh was included). At the corps levels, the range of relative percentage differences was 0.86 percent for IV corps to 2.43 percent for III corps, when Gia Dinh was included (1.49 percent when excluded).

The province-by-province analyses formed the basis for the study and the evaluation. The average relative percentage differences of 1.4 percent ($\leq$ 2 percent for 12 provinces) involved 52 percent of the VC Controlled, weighted, population. When the average difference was about 2.5 percent ($\leq$ 3 percent for 20 provinces), 76 percent of the VC Controlled, weighted was accounted for. And 97.5 percent of the VC Controlled, weighted was covered when the average relative percentage difference was about 7.7 percent ($\precsim$ 10 percent for 39 provinces). Four provinces, including Gia Dinh, accounted for the largest (over 10 percent) differences, and related to about 2.5 percent of VC Controlled, weighted; Gia Dinh accounted for about half of the 2.5 percent.

The results should not, of course, be assessed simply in terms of the percentage differences. The statistical and numerical results for all the provinces taken together are judged to merit a conclusion that the investigation was a success, and certain observations concerning individual provinces further reinforce this conclusion. For example, the 26 provinces that measured progress (i.e., VC Controlled, weighted) with a relative percentage difference of equal to or less than 5 percent (or an average of 3.6 percent) included 84 percent of the estimated VC population. At the same time, 17 provinces exhibited less satisfactory results. Some explanations are in order.

Based on knowledge, which could not become directly a part of the quantitative measurements, that certain data deficiencies existed (for both the dependent and independent variables), the 17 provinces were examined separately.

(C)

TABLE D1

Comparative Results between Adjusted Reported Data and Computed Data for All Provinces

(May 1964–February 1966)

(1) Key number[a]	(2) Province	(3) Corps	(4) Approximate average total population, thous	(5) Range of VC Controlled, weighted, %	(6) Average VC Controlled, weighted, %	(7)[b] Standard error of average proportion of VC Controlled, weighted, %	(8)[c] Average population in VC Controlled, weighted, thous	(9)[d] Standard error, population in VC Controlled, weighted, thous	(10)[e] Relative error in VC Controlled, weighted, %	(11) Explanation of data deficiency
1	Quang Tri	I	275	36–48	43.6	1.55	120.0	4.2	3.5	—
2	Thua Thien	I	460	28–36	32.7	0.82	150.6	3.8	2.5	—
3	Quang Nam	I	565	50–78	67.9	1.08	383.7	6.1	1.5	—
4	Quang Tin	I	355	62–72	67.0	1.39	237.8	4.9	2.0	—
5	Quang Ngai	I	650	47–55	52.5	1.41	341.6	9.1	2.6	—
6	Binh Dinh	II	850	61–67	65.2	1.01	554.3	8.5	1.5	—
7	Phu Bon	II	45	30–52	39.6	3.97	17.8	1.7	10.0	f
8	Pleiku	II	160	41–69	55.3	3.99	88.5	6.3	7.2	f
9	Kontum	II	105	27–51	40.2	1.41	42.2	1.4	3.5	—
10	Quang Duc	II	30	11–30	21.5	2.21	6.4	0.6	10.3	f,g
11	Darlac	II	200	50–58	54.8	1.33	109.7	2.6	2.4	—
12	Phu Yen	II	340	62–71	66.0	1.99	224.2	6.7	3.0	—
13	Khanh Hoa	II	295	14–30	21.6	1.26	63.7	3.7	5.8	g
14	Tuyen Duc	II	80	2–17	8.3	2.55	6.6	2.0	30.5	f,g
15	Ninh Thuan	II	145	7–22	12.6	4.64	18.2	6.7	36.8	f,g
16	Lam Dong	II	65	18–29	23.4	2.22	15.2	1.4	9.5	f,g
17	Binh Thuan	II	250	24–33	30.1	0.73	75.4	1.8	2.4	—
18	Binh Tuy	III	63	29–40	32.6	1.67	20.5	1.0	5.1	f
19	Binh Long	III	80	41–54	48.6	2.69	38.9	2.1	5.5	f
20	Phuoc Long	III	65	38–76	56.6	4.71	36.7	3.0	8.3	f,h,i
21	Phuoc Thanh[j]	III	—	—	—	—	—	—	—	h
22	Long Khanh	III	125	15–22	18.6	1.07	23.2	1.3	5.7	g,h,i
23	Bien Hoa	III	310	13–32	22.1	2.27	68.5	7.0	10.2	g,h,k
24	Phuoc Tuy	III	100	30–37	24.2	1.08	24.2	1.0	3.1	—
25	Long An	III	380	60–70	64.2	0.86	224.2	3.3	1.3	—
26	Hau Nghia	III	225	69–74	72.1	1.19	162.3	2.6	1.6	—
27	Tay Ninh	III	235	23–64	43.6	3.42	102.5	8.0	7.8	i
28	Binh Duong	III	240	47–56	52.8	1.22	126.6	2.9	2.3	h
29	Dinh Tuong	IV	520	52–65	58.2	0.97	302.9	5.0	1.6	—
30	Go Cong	IV	170	47–60	54.9	0.53	93.2	0.9	0.9	—
31	Kien Hoa	IV	540	52–61	58.0	0.99	313.5	5.3	1.7	—
32	Kien Tuong	IV	58	37–47	41.8	1.40	24.2	0.8	3.3	—
33	An Giang	IV	470	0–22	10.5	1.17	49.6	5.5	11.1	g
34	Chau Doc	IV	445	16–40	32.1	2.00	142.8	8.9	6.2	h
35	Kien Phong	IV	290	22–48	36.2	1.70	104.9	4.9	4.7	—
36	Vinh Binh	IV	540	47–55	51.7	0.71	279.2	3.8	1.3	—

37	Vinh Long	IV	555	22–30	27.8	0.81	154.4	4.5	2.9	—
38	An Xuyen	IV	225	67–71	69.7	0.52	156.8	1.1	0.7	—
39	Ba Xuyen	IV	360	33–52	43.3	3.15	155.9	11.3	7.2	h
40	Bac Lieu	IV	260	53–56	52.7	0.78	137.1	2.0	1.4	—
41	Chuong Thien	IV	250	55–60	58.7	0.77	146.7	1.9	1.3	—
42	Kien Giang	IV	365	33–42	39.2	1.47	143.0	5.3	3.7	—
43	Phong Dinh	IV	420	40–53	46.6	1.33	195.7	5.5	2.8	—
44	Gia Dinh (CMD)	III	965	0.2–12	7.2	1.91	69.8	18.4	26.4	g,k
	Summary		13,126	0–78	44.0	—	5775.4	—	—	—

[a]As on frontispiece.
[b]Square root of variances of proportion of population in VC Controlled, weighted; values of variances omitted.
[c]Col. 4 times col 6.
[d]Col 4 times col 7.
[e]Col 9 divided by col 8.
[f]Low population density; mainly highland and Montagnard.
[g]Small average percentages reported generate larger percentage differences.
[h]Province added, deleted, or combined.
[i]Large VC force concentration.
[j]Deleted; divided among 20, 22, 23, and 28 by administrative boundary changes.
[k]Large RVN force concentration; effect on population control not adequately resolved.

(C)

TABLE D2

Summary of Comparative Results between Adjusted Reported Data and Computed Data for All Provinces

(May 1964–February 1966)

Intervals, percentage relative error, computed vs reported	Average, percentage relative error, computed vs reported	Provinces	Provinces, % of total	Provinces, cumulative % of total	Percentage of VC Controlled, weighted, population	Cumulative percentage of VC Controlled, weighted, population
≤ 2	1.4	12	28	28	52.0	52.0
≤ 3	2.5	8	19	47	24.0	76.0
≤ 5	3.6	6	14	61	8.0	84.0
$\overset{\sim}{<} 10$	7.7	13	30	91	13.5	97.5
> 10	26.2	4	9	100	2.5	100.0

(C)

TABLE D3

Comparative Results between Adjusted Reported Data and Computed Data, Countrywide and Corps

(May 1964–February 1966)

Area	Approximate average total population, thous	VC Controlled, weighted		Percentage relative error, VC Controlled, weighted, population, thous	Percentage relative error, VC Controlled, weighted
		Average population, thous	Percent controlled		
Countrywide	13126	5775.3	44.0	31.7	0.55
Countrywide, Gia Dinh excluded	12161	5705.5	46.9	34.4	0.60
I Corps	2305	1233.9	53.5	13.4	1.08
II Corps	2565	1222.8	47.7	15.6	1.27
III Corps	2788	917.9	32.9	12.6	1.49
III Corps, Gia Dinh excluded[a]	1823	848.1	46.5	22.4	2.43
IV Corps	5468	2400.7	43.9	20.6	0.86

[a]Large RVN force concentration in Capital Military District; effect on population control not adequately resolved.

Note that direct access to the sources providing the data could have greatly improved the results; i.e., data deficiencies, when detected, could have been at least partly corrected. Equally important, any extension of the investigation could profit from the experience gained.

The most probable reasons for data deficiencies, and hence less accurate computations, are shown in Table D1 and listed as follows:

(a) Geographical and demographic; i.e., relatively small province populations and very low population densities resulted in relatively poor assessments. Also, enemy intelligence, especially on VC main and local forces, would be affected. For example, 9 of the 17 provinces, largely highland and Montagnard regions, fall into this category (frontispiece, Key Nos. 7, 8, 10, 14, 15, 16, 18, 19, and 20). All 17 have at least one associated, probable reason; some have more than one.

(b) Provinces reporting relatively small average percentages (less than 25 percent) of VC Control can more readily generate larger percentage differences (frontispiece, Key Nos. 10, 13, 14, 15, 16, 22, 23, 24, 33, and 44).

(c) At intervals during the test period, new provinces were formed out of one or more adjacent provinces, or deleted and combined with others. As a result the reporting system was upset for a time and required changes and redistributions in the data bases in order to present a continuous record (frontispiece, Key Nos. 20, 21-deleted, 22, 23, 28, 34, and 39).

(d) Several provinces with relatively large VC force concentrations, as those including or nearby Zones C and D, were affected (frontispiece, Key Nos. 20, 22, and 27).

(e) Gia Dinh (44), also designated the CMD, especially, and to some extent Bien Hoa (23), presented a special problem. Large numbers of RVN and other regular forces were based in these provinces. What proportion of these forces should be taken (as an independent variable) as affecting population control in the computations was apparently not adequately resolved.

REFERENCES

CITED REFERENCES

1. Dept of Army, Deputy Chief of Staff for Operations, Reporting System for Population Control, May 1964.
2. D. B. Duncan and M. Walser, "Multiple Regression Combining Within- and Between-Plot Information," Biometrics, 22 (1): 26–43 (1966).
3. Library of Congress, Photoduplication Service, "FORTRAN II Program: Suitable for IBM 7094," Document No. 8779, Washington, D. C., Mar 66.

ADDITIONAL REFERENCES

Cramer, H., Mathematical Methods of Statistics, Princeton University Press, Princeton, N. J., 1951.

Ezekiel, M., and K. A. Fox, Methods of Correlation and Regression Analysis, John Wiley & Sons, Inc., New York, 1959.

Johnston, J., Econometric Methods, McGraw-Hill Book Company, Inc., New York, 1960.

Kendall, M. G., and A. Stuart, The Advanced Theory of Statistics, Vol 2, Hafner Publishing Co., New York, 1961.

Kendall, M. G., A Course in Multivariate Analysis, Griffin's Statistical Monographs & Courses, Charles Griffin & Company, Ltd., London, 1961.

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Research Analysis Corporation McLean, Virginia 22101	2a. REPORT SECURITY CLASSIFICATION CONFIDENTIAL 2b. GROUP 4

3. REPORT TITLE

MEASUREMENT OF PACIFICATION PROGRESS IN VIETNAM (U)

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical Paper

5. AUTHOR(S) *(First name, middle initial, last name)*

James W. Johnson
Charles Anello

6. REPORT DATE September 1968	7a. TOTAL NO. OF PAGES 76	7b. NO. OF REFS 8

8a. CONTRACT OR GRANT NO. DA 44-188-ARO-1 b. PROJECT NO. 008.106 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) RAC-TP-309 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT In addition to security requirements which apply to this document and must be met, each transmittal outside the Department of Defense must have prior approval of Office of Chief of Research and Development, ATTN: STA Division, Department of the Army, Washington, D.C. 20310.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Department of Army, Institutional Research

13. ABSTRACT (U) The purpose of the study was to develop and test a mathematical model, employing the standard linear multiple regression method, to determine whether quantitative measurements of progress in Vietnam, based on data generated in the field, could be made. A satisfactory model would help also to identify and measure factors affecting progress, provide a useful predictive device, and offer a method for checking the field reporting systems. Progress was measured by the proportion of the population under the control of both sides, changes, and the rate of change. Work covered a 22-month test period, May 1964—February 1966, for all 43 provinces; the province data were combined to produce corps tactical zone and countrywide assessments. Statistical tests of the method were satisfactory. A comparison between the reported and computed data was made; at province level percentage differences were less than 5 percent in the majority of cases; at corps level, less than 2 percent; and countrywide, less than 1 percent. The multiple regression calculations were completely computerized, offering rapid assessments. The quality of the results indicated that the technique could help to determine progress, estimate force requirements, and to predict future trends.

DD FORM 1 NOV 65 1473

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Insurgency						
Counterinsurgency						
Vietnam						
Pacification progress, measurement						
Population control						
Mathematical model						
Linear multiple regression						
Factors affecting progress						
Force capabilities and requirements						
Prediction of trends						
Tests of field reporting systems						
Computerization of mathematical models						

[redacted]

PRO CON

RAC

[redacted]